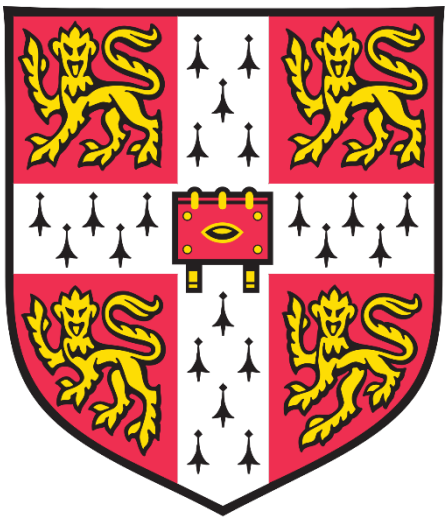
Elucidating the Genetic Basis of Canine Progressive Retinal Atrophies in Several Breeds of Dog



Rebekkah Jean Hitti-Malin

Department of Veterinary Medicine

Hughes Hall College

University of Cambridge

This thesis is submitted for the degree of

Doctor of Philosophy

May 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Signed: _____

Date: 4th May 2020

Rebekkah Jean Hitti-Malin

Elucidating the Genetic Basis of Canine Progressive Retinal Atrophies in Several Breeds of Dog

Rebekkah Hitti-Malin

Canine progressive retinal atrophies (PRA) are a group of hereditary diseases characterised by rod and cone photoreceptor cell death in the retina. This study sought to elucidate novel PRA-associated variants of distinct forms of PRA in three breeds of dog: the Lhasa Apso (LA), Giant Schnauzer (GS) and Shetland Sheepdog (SS).

A genome-wide association study identified a 1.3 Mb disease-associated region on canine chromosome 33 in LA dogs. Whole genome sequencing (WGS) analysis of a PRA-affected LA revealed a long interspersed nuclear element-1 (LINE-1) insertion in the predicted promoter region of the retinal candidate gene, *IMPG2*. Validation of the LINE-1 insertion determined it segregated with disease and was likely to be private to LA dogs.

Comprehensive WGS analyses alone were utilised to determine PRA-associated mutations in a family of GS dogs, and in a single SS. For the GS breed, WGS was performed on two affected siblings and both non-affected parents. Successive filtering, for autosomal recessive deleterious variants, against 568 canine genomes identified a single nucleotide variant (SNV) in the gene encoding NECAP endocytosis associated 1 (*NECAP1*): c.544G>A (p.G182R). Screening 5,130 canids revealed only the three PRA-affected GS were homozygous for the SNV, yet heterozygotes were identified in the GS breed and in other breeds of German ancestry. *NECAP1* has not previously been associated with retinal degeneration; however, these findings, in parallel with known gene function, indicate *NECAP1* should be considered as a strong candidate for retinal degeneration research in other species. Following WGS analysis of the single SS against 176 controls of other breeds, a c.1222G>C (p.A408P) SNV in the Bardet-Biedl syndrome 2 gene (*BBS2*) was identified. In addition to PRA, homozygotes exhibited features including an upturned nose, unusual coat and dental defects, proposing a novel syndromic form of canine PRA.

This research has elucidated three novel PRA-associated mutations for which diagnostic DNA tests have been developed, offering breeders the opportunity to avoid producing PRA-affected dogs. Since PRA shares clinical features to human retinitis pigmentosa (RP) and other phenotypically similar retinal diseases, this study may offer novel insights for consideration in human as well as canine retinal degeneration research and gene therapies.

Acknowledgements

Firstly, my thanks go to my supervisors for making this thesis possible. Thank you to Dr Cathryn Mellersh for her outstanding support and guidance throughout my time at the Animal Health Trust. From the start of my career, she has been encouraging and patient in helping me achieve and progress, and has provided substantial advice and expertise. It has been a pleasure working with Cathryn, who has hugely impacted my scientific development and has made this PhD experience extremely rewarding and fulfilling. Thank you to Dr David Sargan, my supervisor at Cambridge, who has also provided immense support and advice throughout my PhD. David has provided me with a wealth of knowledge, and I am most grateful for his time he has dedicated to teaching and encouraging me to progress throughout this PhD.

I would also like to thank my colleagues and friends in the Canine Genetics team at the Animal Health Trust, especially Dr Louise Burmeister and Dr Sally Ricketts for their kindness, advice and proofreading skills not only throughout this PhD, but in all aspects during my time at the Animal Health Trust. I am very grateful for their expertise and knowledge. Thanks also go to Dr Debbie Guest, my PhD mentor - our final year chats gave me confidence to explore new opportunities following my PhD studies. I am also hugely grateful to Debbie for her scientific expertise in cell techniques to help me tackle the luciferase assay experiments.

Thank you to all the funders who supported this work financially, in particular the Animal Health Trust, the Kennel Club Charitable Trust and all dog owners, vets and breed clubs for their financial donations and submission of valuable research samples and phenotype information that enabled these studies to build a thesis.

Last, but certainly not least, thank you to my incredible network of family and friends that have supported me throughout my PhD studies. To my Harley, my loving husband - you have always believed in me. Thank you for your encouragement and for keeping me grounded and sane throughout this PhD. To my parents, Nanny and sisters; thank you for your continuous love, help and support in whatever I pursue.

Table of Contents

Abbreviations

×g	Times gravity
-log_{10}	Negative log_{10}
ΔRn	Delta Rn
A	Adenine
AAV	Adeno-Associated Virus
AFLP	Amplified Fragment Length Polymorphism
AgRP	Agouti-related peptide
AHT	Animal Health Trust
AP-	Adaptor protein
bp	Base pair
BAM	Binary Alignment/Map
BBS	Bardet-Biedl syndrome
BBS2	Bardet-Biedl syndrome 2
BVA/KC/ISDS	British Veterinary Association/Kennel Club/International Sheep Dog Society
BWA	Burrows-Wheeler-Alignment
C	Cytosine
Ca	Calcium
CANFA	Canis familiaris chromosome
CCT	Chaperonin containing T-complex protein-1
CCVs	Clathrin-coated vesicles
cDNA	Complementary DNA
CDR	Done-rod degeneration
cGMP	Cyclic guanosine monophosphate
CME	Clathrin mediated endocytosis
CMV	Cytomegalovirus
CNG	Cyclic nucleotide gated
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/Cas9

CRX	Cone-rod homeobox
Ct	Cycle threshold
DBVDC	Dog Biomedical Variant Database Consortium
ddH_2O	Ultrapure water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EtBr	Ethidium bromide
EDTA	Ethylenediaminetetraacetic acid
EIEE	Early infantile epileptic encephalopathy
EMMAX	Efficient Mixed-Model Association eXpedited
EOSRD	Early-onset severe retinal dystrophy
ERD	Early retinal degeneration
ERG	Electroretinogram
FID	Family ID
g	Gram
G	Guanine
GATK	Genome Analysis Toolkit
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GS	Giant Schnauzer
GTP	Guanosine triphosphate
gVCF	Genomic Variant Call Format
GWAS	Genome-wide association study
HC	Hereditary cataracts
HEK239T	Human embryonic kidney 293T
ID	Identity
IFT	Intraflagellar transport
IGV	Integrative Genomics Viewer
IID	Individual identity

IMPG2	Interphotoreceptor matrix proteoglycan 2
INDEL	Insertions and deletions
IPM	Interphotoreceptor matrix
IRD	Inherited retinal degeneration
K	Potassium
Kb	Kilobase
L	Litre
LA	Lhasa Apso
LCA	Leber congenital amaurosis
LD	linkage disequilibrium
LINE-1	Long interspersed nuclear element-1
M	Molar
MAF	Minor allele frequency
Mb	Megabase
MDCK	Madin-Darby canine kidney
MDS	Multi-dimensional scaling
MSC	Mesenchymal stem cell
mIL-17	Mouse interleukin-17
N	Normality
Na	Sodium
NaOH	Sodium hydroxide
NECAP1	NECAP endocytosis associated 1
NGS	Next-generation sequencing
NTC	No template control
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDE	Phosphodiesterase enzyme
PIREs	Pineal regulatory elements

POMC	Proopiomelanocortin
PRA	Progressive retinal atrophy
PRCD	Progressive rod-cone degeneration
QC	Quality Control
qPCR	Quantitative PCR
Q-Q	Quantile-Quantile
R^2	Coefficient of correlation
RCD	Rod-cone degeneration
RetNet	Retinal Information Network
RHO	Rhodopsin
RLU	Relative light units
Rn	Normalised reporter
RNA-seq	RNA sequencing
ROH	Runs of homozygosity
RP	Retinitis pigmentosa
RPE	Retinal pigment epithelium
RPGRIP1	Retinitis pigmentosa GTPase regulator-interacting protein 1
SAM	Sequence Alignment/Map
SARD	Sudden acquired retinal degeneration
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SS	Shetland Sheepdog
SV40	Simian virus 40
T	Thymine
TBP	TATA box binding protein
TFBS	Transcription factor binding site
UTR	Untranslated region
VCF	Variant Call Format
VEP	Variant Effect Predictor

WES	Whole exome sequencing
WGS	Whole genome sequencing
α	Alpha
λ	Lambda – inflation factor
Metric prefixes	
p	Pico (10^{-12})
n	Nano (10^{-9})
µ	Micro (10^{-6})
m	Milli (10^{-3})
K	Kilo (10^{3})
M	Mega (10^{6})
G	Giga (10^{9})

List of Figures

List of Tables

Chapter 1 Introduction

1.1 The origin of domestic dog breeds

Throughout history, humans have encouraged a mutually beneficial relationship with the dog (*Canis lupus familiaris*), leading to its domestication. Phenotypic diversity of domestic dogs suggests a varied genetic heritage, which may have been influenced by interbreeding with wild wolf populations (Vila et al, 1997). Ancestry of the modern dog is predominantly European and insights into the history of dog domestication in Europe suggest the modern dog population arose from a single domestication event occurring between 20,000 - 40,000 years ago from a population of grey wolves (Botigué et al, 2017; Parker et al, 2004). Two major bottleneck events throughout history are considered to have influenced genetic diversity in the dog genome. The first population bottleneck shaped the ancient domestication of the dog from the grey wolf (Savolainen et al, 2002). The formation of dog breeds over 200 years ago resulted in a second bottleneck when dogs were selectively bred for certain purposes, such as herding and hunting, and for desirable characteristics and traits including size, shape, coat colour, temperament, and obedience. Although individual breeds are genetically distinct populations, they appear to have diverged simultaneously (Parker et al, 2004). Today, 352 dog breeds are recognised worldwide (Fédération Cynologique Internationale, 2020) with over 200 of these different dog breeds recognised by The Kennel Club UK, the largest registration body for dogs in the UK. The population of purebred dogs in the UK is ever growing, with over 250,600 dogs registered by The Kennel Club in 2018, an increase of 7,300 puppies from 2017 (The Kennel Club UK, 2019a). This number accounts for a small portion of the dog population in the UK, with a total of 8.9 million dogs estimated to be owned across the UK in 2018 (PFMA, 2019).

1.2 The canine genome

Linkage maps of the canine genome (Lingaas et al, 1997; Mellersh et al, 1997; Neff et al, 1999) provided a foundation for the development of a complete chromosome map for the dog in 1999, comprising 38 autosomes and the sex chromosomes (Breen et al, 1999; Yang et al, 1999). The first high-quality draft (7.5X) canine genome sequence was published in 2005 using a Boxer dog as the reference genome, and is approximately 2.4 gigabases (Gb) in size containing approximately 19,000 genes, with nearly all genes homologous to known human genes (Lindblad-Toh et al, 2005). With this, a dense map of single nucleotide polymorphisms (SNPs) across breeds were also reported. The paleogenomic study of prehistoric dogs provides evidence that the recent canine reference genome, CanFam3.1 (Sept.2011.Broad CanFam3.1/canfam3), shares over 60% of its identity to 7,000 year old dog specimens, dated back to Early Neolithic times (Botigué et al, 2017). Historical dog domestication and breed creation shaped the haplotype structure of canine breeds,

which is reflected in the canine genome where interbreed variation is present. A genetic study of 85 dog breeds demonstrated that humans and dogs share similar levels of overall nucleotide diversity: 8×10^{-4} nucleotide substitutions per base/pair. However, the genetic variation between dog breeds (27.5%) is much greater than the variation between human populations (5.4%); as is the degree of genetic homogeneity (94.6% in dog breed populations versus 72.5% in human populations) (Parker et al, 2004). As a result, high levels of linkage disequilibrium (LD; non-random association of alleles at different loci) are present within dog breeds, extending over several megabases (Mb), in comparison to LD across breeds, which extend over tens of kilobases (Kb) (Lindblad-Toh et al, 2005). The unique pattern of LD in dogs provides the opportunity to study the molecular basis of inherited traits, facilitating the rapid identification of loci associated with complex and Mendelian traits and diseases, relevant to dogs and other species.

1.3 Genetic implications of intensive breeding

The dog plays a key role in human lives; however, its domestication and integration into society has driven selection for breed-related characteristics. Intensive selection for desired physical or behavioural traits and the introduction of closed stud books led to the generation of isolated breeding populations, each with a limited number of founders with reduced phenotypic and genotypic diversity. The same selective pressures have led to long stretches of DNA in LD (Goldstein et al, 2006; Sutter et al, 2004), where inter-breed LD is shorter than LD within breeds. As a consequence, there is little genetic variability within breeds, with an increase in homozygosity, and relatively large genetic differences between breeds. High levels of homozygosity throughout the genome increases the likelihood that a variant that arose spontaneously in earlier generations occurs within a region of homozygosity in the current population. Breeding closely related dogs that share similar chromosomal DNA sequences harbouring trait- and disease-associated variants leads to an increased incidence of inherited diseases. Thus, by selecting for a given trait, undesirable disease-causing mutations can also be unintentionally inherited.

Animals with particularly favourable characteristics are often bred repeatedly, where popular males in the dog population are often used extensively, commonly known as the popular sire effect. This is a problematic breeding practice as extensively breeding with one sire limits the gene pool, thereby reducing genetic diversity resulting in an over representation of allelic pools in subsequent generations (Parker et al, 2004; Quignon et al, 2007). To maintain favourable traits in future offspring, breeders may choose not only to continue to breed with one popular sire but also with its close relatives. This, combined with historical genetic bottlenecks and breeding in closed populations, increases the incidence of inherited disease in offspring (Mellersh, 2008). Breeding of closely related individuals (inbreeding) can cause inbreeding depression, which can incur a deleterious effect on a given trait and subsequently increases disease occurrence. In dogs, this is

demonstrated in a study in Standard Poodles, where an increased incidence of two autoimmune disorders in the breed probably arose due to an artificial genetic bottleneck and inbreeding (Pedersen et al, 2015). This suggests that generally a reduction in genetic diversity within a breed has a negative impact on canine health, increasing the incidence of deleterious complex disorders and recessive traits, including heritable eye conditions.

1.4 The canine eye

The eye is a complex organ with specialised structures and functions to provide optimal vision. Ocular development is a multifaceted process with various crucial stages, pathways and signalling interactions involved. Across mammals, anatomy and function of the eye varies depending on species evolution and adaptation in different environments. The anatomy of the canine eye (Figure 1.1) and human eye have many similarities, sharing most of their anatomical features. One notable difference is the existence of a layer of reflective tissue called the tapetum lucidum in the canine eye. This layer serves to reflect light after it has passed through the retina, re-stimulating photoreceptor cells of the retina and causing "eyeshine" at night when illuminated, to aid night vision. The tapetum lucidum occupies approximately half of the canine fundus and shows considerable variation across individuals in terms of its colouration, shape and contour, with both age and breed variations (Jacobs et al, 1993). The fundus is the posterior inside of the eye housing the retinal neural tissue, upon which ophthalmoscopic examination can reveal retinal abnormalities associated with disease.

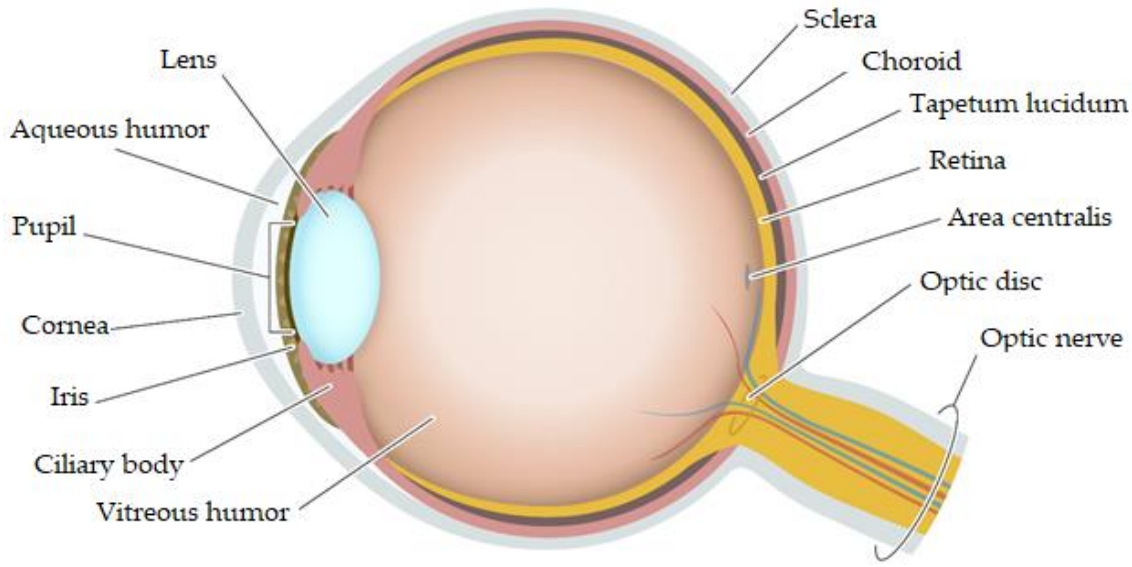


Figure 1.1: Schematic diagram demonstrating the anatomy of the canine eye. (Adapted from CleverPet).

Most of the canine retina is dominated by rod photoreceptor cells with an uneven distribution of cone photoreceptors. The fovea, an area with dense cone photoreceptors, is present in the retinas

of many primates and enables high resolution vision and acuity. Historically it was believed that canids lack this region, where the fovea is replaced by the 'area centralis' within a visual streak: a horizontal distribution of high retinal ganglion cell density in the retina (Kemp & Jacobson, 1992). This area centralis is located supero-temporal to the optic nerve head and is characterised by a high concentration of retinal ganglion cells (Peichl, 1992a). It is believed the visual streak has evolved from the wolf retina and provides high visual acuity when looking to the horizon (Peichl, 1992b). A more recent study has shown the area centralis in the canine retina anatomically resembles the primate fovea composed exclusively of cones (Beltran et al, 2014), however, it is unclear if this is seen across all dogs or if this is a recent adaptation in the formation of domestic breeds. Despite limited studies to define the extent of the area centralis in domestic dogs, there is evidence that the visual streak and area centralis vary independently amongst breeds, with variation in retinal ganglion cell density and distribution across dog breeds. A study by McGreevy et al. (2004) found a negative correlation between the ratio of peak ganglion cell density in the area centralis to visual streak and skull length in dogs. This suggests that selective breeding for certain skull shapes in dog breeds has resulted in variability in the area centralis in the canine retina.

1.5 The retina

Forming the inner lining of the posterior globe, the retina is a specialised tissue of the central nervous system. In order to transmit signals to the brain via the optic nerve, the retina converts light stimuli received through the lens to electrical potentials. This enables information to be interpreted as vision. The retina is a multi-layered structure consisting of neural layers, and a layer of photoreceptor cells supported by a non-neural layer of pigmented epithelial cells, termed the retinal pigment epithelium (RPE) (Figure 1.2).

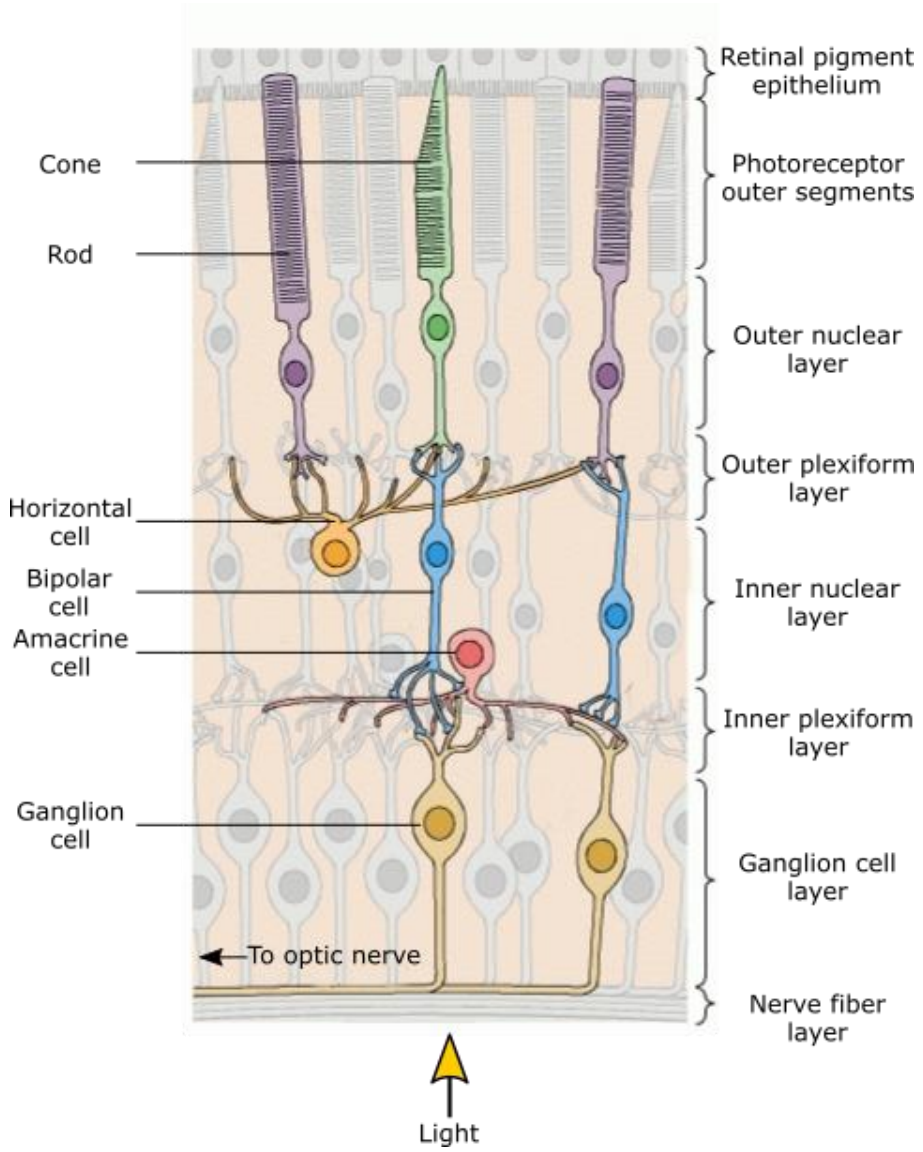


Figure 1.2: Diagram of the layers of the retina (adapted from Purves, 2001).

Retinal pigment epithelium

The RPE is a melanin-rich structure situated between the neuroretina and the underlying vascular layer known as the choroid. The apical membrane of the RPE faces the outer segments of specialised neuron cells called photoreceptors, and its basolateral membrane facing Bruch's membrane (reviewed in Simó et al, 2010), an extracellular matrix located at the innermost layer of the choroid. The RPE has many functions essential to maintain retinal integrity. This includes the transportation of nutrients from the blood to photoreceptors while removing ions, metabolic end products and water from the retina to the blood. The RPE also mediates the visual cycle, absorbing excess light entering the eye to protect against photooxidation as well as carrying out essential reactions in phototransduction including the re-isomerisation of all-*trans*-retinal into 11-*cis*-retinal (reviewed in Bibliowicz et al, 2011). The RPE also plays a crucial role in supporting and nourishing photoreceptor cells, including the phagocytosis of shed photoreceptor outer

segments, and stabilizing ion composition within the subretinal space to contribute to maintaining photoreceptor excitability (reviewed in Simó et al, 2010). Surrounding the RPE and the photoreceptor outer segments is the interphotoreceptor matrix (IPM), a unique extracellular complex integral to proper function of retinal photoreceptors (Kuehn et al, 2001; Lazarus & Hageman, 1992). The IPM aids in recycling photoreceptor outer segments; facilitates retina-RPE adhesion; establishes a suitable milieu for photoreceptor survival; and facilitates the exchange of molecular products between the RPE and photoreceptor cells (Acharya et al, 2000; Kuehn & Hageman, 1999; Lazarus & Hageman, 1992).

Photoreceptors

Two types of photoreceptors are present in the retina, differing in morphology and function: rods and cones. Rod and cone photoreceptors are specialised elongated cells that are designed to capture and convert light into an electrical signal using specialised ciliary organelles to detect light. These organelles, termed photoreceptor outer segments, are derived from microtubule-based organelles on the cell surface called the primary non-motile cilium: tiny hair-like organelles attached to the cell surface involved in signal transduction from the environment or from other cells (reviewed in Purves, 2001).

Rod and cone photoreceptors provide different visual functions and are named after their shapes (Figure 1.3). Rod outer segments are cylindrically shaped, function in dim light and lack colour perception. Cone outer segments are conical and work in bright light, enabling high-acuity and colour vision. Each type also differs in the type of photopigment they contain, arrangement of membranous discs, distribution across the retina and pattern of synaptic connections (reviewed in Purves, 2001). The photoreceptor layer is dense with rod and cone outer segments closely arranged in stacks of flattened membranous discs, in which visual pigments are embedded. Rhodopsin, the light-sensitive photopigment present in rod disc membranes, is comprised of the light-absorbing molecule 11-*cis*-retinal bound to a lipoprotein component called 'opsin'. In humans, there are three subclasses of cone photoreceptors for trichromatic colour vision, long- (L), medium- (M) or short-wavelength-sensitive (S) cones, each with one of three photopigments (red, green or blue, respectively) present in disc membranes. Dogs have dichromatic vision to perceive colour with the presence of two subclasses of cone photoreceptors in the canine retina, each with its own specific red/green or blue pigment to form visual pigments, collectively called 'photopsins'. These are known as long- and medium-wavelength-absorbing cones (L/M cones) with a maximum sensitivity of 555 nanometre (nm) and short-wavelength-absorbing cones (S cones) of 429 nm sensitivity (Jacobs et al, 1993; Neitz et al, 1989).

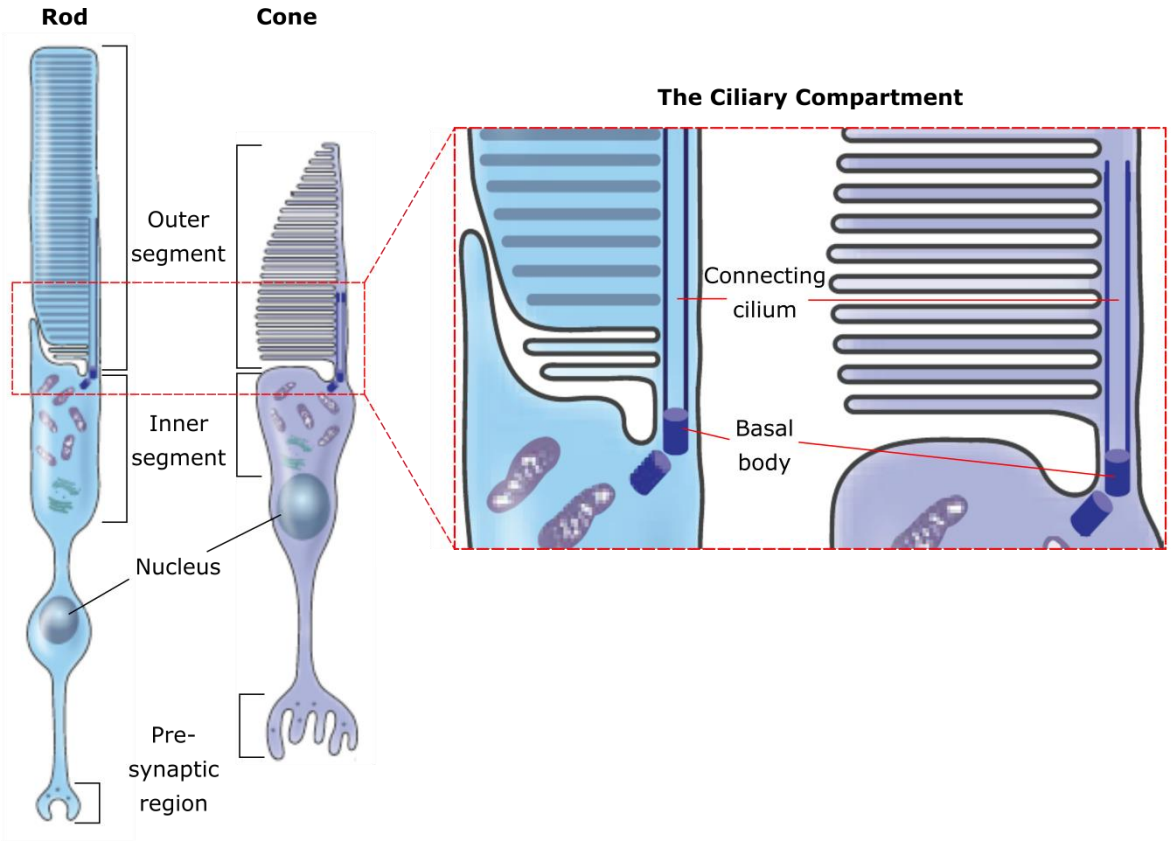


Figure 1.3: Schematic diagram of a rod and cone photoreceptor. The outer segment includes hundreds of tightly packed membranous discs containing visual pigment specific to each photoreceptor cell. The inner segment contains the cell's metabolic machinery including subcellular organelles such as mitochondria. The ciliary compartment is made up of three subunits: the basal body, connecting cilium and the outer segment. The connecting cilium transports proteins specific to the visual system to the outer segment (adapted from Veleri et al, 2015).

Rod and cone photoreceptors also differ in phototransduction mechanisms in the phototransduction cascade: a series of biochemical pathways in response to light. For example, rods respond to a single light-energised photon whereas over 100 photons are required for cones to produce a comparable response. This is largely due to differences in the current a single photon produces in each receptor type (reviewed in Purves, 2001). Furthermore, the adaptation mechanisms of cones are more effective than rods, where even though both adapt to operate across a range of illumination levels, cones recover four times faster than rods when exposed to bright conditions that produce the maximum change in photoreceptor current (reviewed in Purves, 2001).

Due to its ciliary origin, the photoreceptor outer segment shares basic structural elements and common mechanisms for development and maintenance with primary cilia, however there are many features that are unique to the outer segment. The outer segment is more elongated than primary cilia with a larger ciliary plasma membrane for increased light sensitivity. Furthermore, unlike primary cilia, the outer segment undergoes continual renewal, with new discs formed at

the base of the outer segment and discs at the distal tip undergo shedding and phagocytosis by RPE cells (reviewed in Goldberg et al, 2016).

The basal body, the connecting cilium and the outer segment are central components located within photoreceptors and are collectively termed the 'ciliary compartment' (Figure 1.3). The localisation of proteins involved in ciliogenesis and trafficking of specific proteins to the outer segment occurs within the ciliary compartment (Seo & Datta, 2017). At the base of the ciliary compartment lies the basal body, playing a crucial role in microtubule organisation. Primary cilia are typically non-motile, with a ciliary axoneme comprising of nine doublet microtubules arising from the triplet microtubules of the basal body within the cell. The ciliary microtubules direct the trafficking of ciliary cargo from the ciliary base to the tip, including vesicles and organelles (reviewed in Mannu, 2014). The targeting and assembly of ciliary cargo is controlled by specialised multiprotein complexes within the cilium, including intraflagellar transport (IFT) proteins and the BBSome: a stable complex consisting of eight BBS proteins (BBS1, 2, 4, 5, 7, 8, 9 and BBS18/BBIP10). The connecting cilium acts as a ciliary transition zone and connects the outer segment to the inner segment of photoreceptors. The inner segment contains subcellular organelles including mitochondria, Golgi apparatus, endoplasmic reticulum and lysosomes, which play a role in the metabolic and biosynthetic mechanisms of the cell. Distal to the inner segment is the nuclear region housing the nucleus, followed by synaptic terminals branching out to interact with horizontal cells, bipolar cells, amacrine cells and ganglion cells across additional retinal layers. The arrangement of these cell types and synapses contribute to the different mechanisms of rod and cones in light conditions to transmit rod and cone information to ganglion cells following phototransduction (reviewed in Purves, 2001).

The phototransduction cascade

Retinal photoreceptors contain the proteins required for the phototransduction cascade: a process where photoreceptors detect photons and convert light stimuli into cellular signals and transmit as a nerve impulse. When light enters the eye, a cascade of chemical conversions and photochemical signalling events occur. Unlike in most sensory systems where an action potential and transmitter release are stimulated, much of the processing within the retina is mediated by graded potentials to initiate a change in the rate of transmitter release onto postsynaptic neurons (reviewed in Purves, 2001).

In darkness, the photoreceptors are active: an abundance of cyclic guanosine monophosphate (cGMP) keeps cGMP-gated channels open, allowing the flow of sodium (Na^+) and calcium (Ca^{2+}) cations into the outer segment, resulting in depolarised photoreceptors that continuously release an excitatory neurotransmitter 'glutamate' from their synapses (reviewed in Purves, 2001). A series of biochemical changes that ultimately lead to a decrease in cGMP levels is initiated when a light-energised photon is absorbed by the photopigment. Rod phototransduction is the most well

characterised of the sensory transduction pathways. When the photopigment absorbs a light-energised photon, it undergoes a conformational change: 11-cis-retinal isomerises to all-trans-retinal and rhodopsin is activated, which in turn catalyses the activation of transducin, a trimeric G-protein, enabling the exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP). This exchange causes the entire alpha (α) subunit of the trimeric transducin molecule to dissociate from the rest of the molecule, enabling it to move and interact with a cGMP phosphodiesterase enzyme (PDE) to hydrolyse cGMP in the disc membrane, producing GMP. This in turn reduces cGMP levels throughout the outer segment and results in fewer cGMP molecules available for binding to the cGMP-gated channels in the outer segment membrane, leading to closure of these channels and a reduction in the influx of cations (reviewed in Purves, 2001). Light exposure therefore leads to a decrease in cGMP levels that initiates closure of the cGMP-gated channels, leaving the cell in a hyperpolarised state (Figure 1.4) (reviewed in Purves, 2001). The Na^+/ Ca^{2+} influx stops, leading to cell hyperpolarisation and termination of glutamate release from the synaptic terminals (reviewed in Gelatt, 2013).

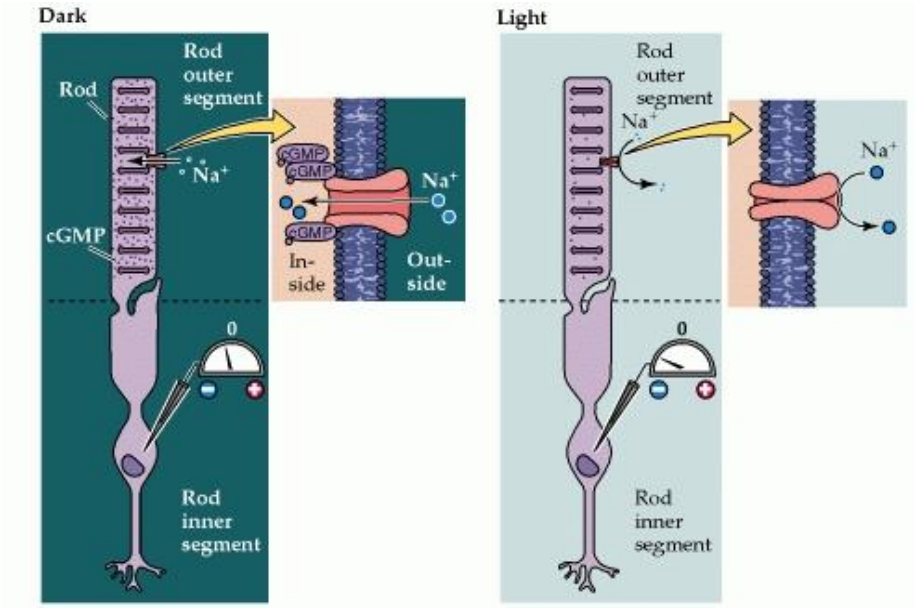


Figure 1.4: An example of light-induced changes in a rod photoreceptor initiated by cyclic GMP-gated channels in the outer segment membrane. cGMP levels in outer segments are increased in the dark and bind to the Na^+ membrane channels. This keeps channels open to allow Na^+ to enter, causing depolarisation of the cell. Light exposure decreases cGMP causing closure of the channels (reviewed in Purves, 2001).

Once a peak light-induced response is reached, the inward current rapidly recovers to the dark state. Inactivation of the cascade involves phosphorylation of rhodopsin by the enzyme 'rhodopsin kinase'. This phosphorylation increases the affinity of the capping protein 'arrestin' to rhodopsin thus blocking further binding and activation of transducin, thus the cascade is unable to continue (reviewed in Mannu, 2014). When GTP is hydrolysed to GDP, PDE and transducin are inactivated

(reviewed in Burns & Arshavsky, 2005). The concentration of Ca^{2+} in the outer segment facilitates light-induced modulation of photoreceptor sensitivity. The cGMP-gated channels in the outer segment are permeable to both Na^{+} and Ca^{2+}, therefore light-induced closure of these channels results in a reduction of Ca^{2+} concentration in the outer segment. This decrease triggers several changes in the phototransduction cascade, all of which tend to reduce the sensitivity of the photoreceptor to light. An example of this are regulatory effects of Ca^{2+} on the cascade, where the reduction in Ca^{2+} concentration in the outer segment results in the activation of retinal guanylate cyclase, which in turn catalyses the regeneration of cGMP. Similarly, the reduction in Ca^{2+} concentration increases the affinity of the cGMP-gated channels for cGMP, reducing the impact of the light-induced reduction of cGMP levels (Purves, 2001). An eventual increase in cGMP concentration causes reopening of the channel and depolarisation of the photoreceptor (reviewed in Gelatt, 2013). Figure 1.5 summarises this phototransduction cascade upon a light-energised response. This sophisticated process highlights the importance of genes encoding photoreceptor proteins and genes involved in phototransduction to convert a light stimulus into vision.

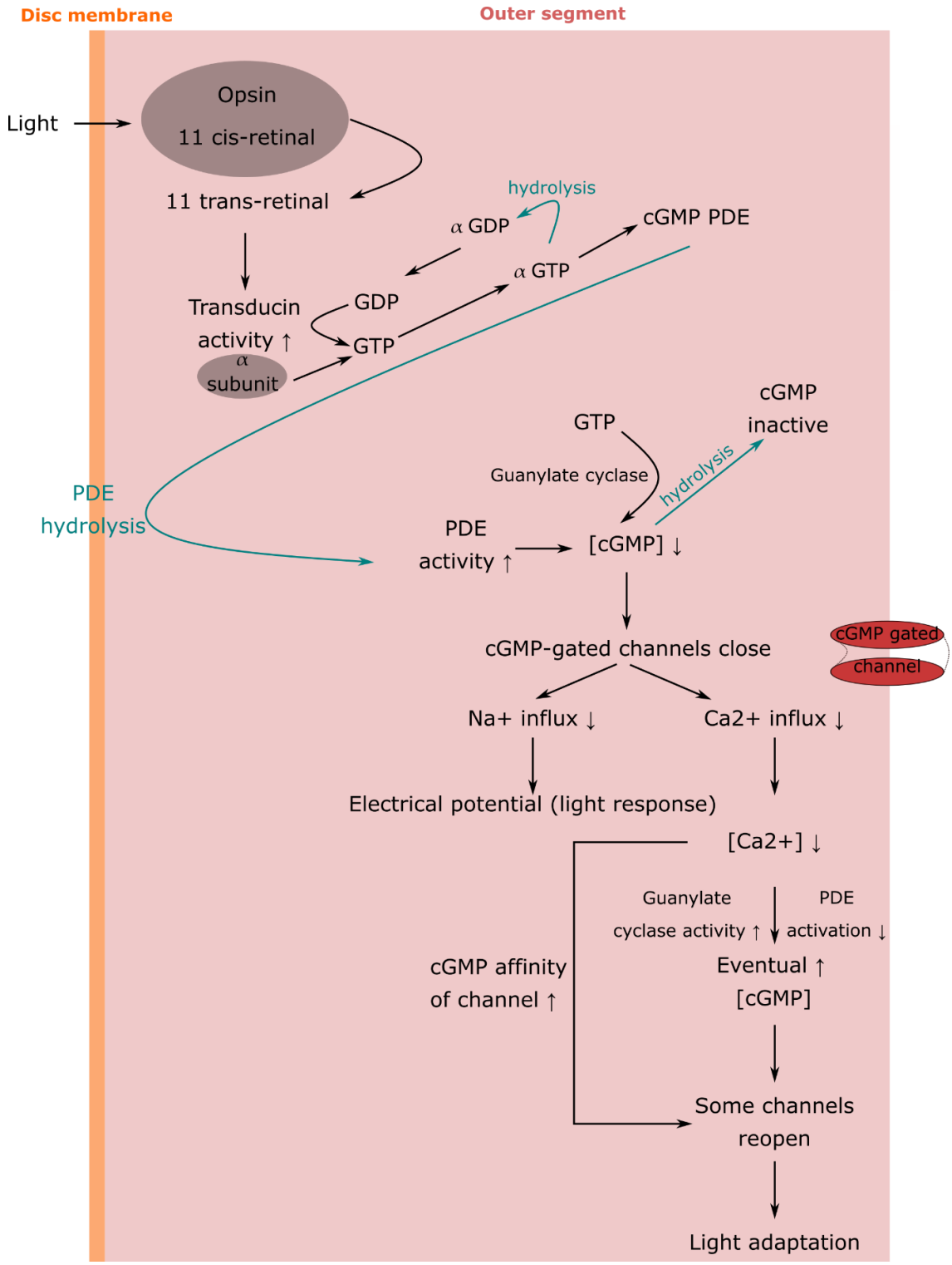


Figure 1.5: Schematic summary of the phototransduction cascade in photoreceptors. Abbreviations include guanosine diphosphate (GDP), guanosine triphosphate (GTP), cGMP (cyclic guanosine monophosphate), PDE (phosphodiesterase enzyme), sodium ions (Na^+), calcium ions (Ca^{2+}), alpha (α) subunit. Square brackets refer to the concentration of a component.

The visual cycle

The phototransduction cascade can only occur if the system can reset itself to allow it to be activated again. Following absorption of a photon through phototransduction, the visual cycle is

required to regenerate 11-*cis*-retinal to restore photosensitivity to rhodopsin. Photoactivated rhodopsin dissociates to form opsin and all-*trans*-retinal (reviewed in Lee et al, 2010). The all-*trans*-retinal generated in response to a light stimulus is transported to the outer segment in the photoreceptor by the retina-specific ATP-binding cassette, subfamily A, member 4 protein (ABCA4) (reviewed in Travis et al, 2007). An enzymatic reduction occurs of all-*trans*-retinal to all-*trans*-retinol (vitamin A) by retinol dehydrogenases. Newly formed all-*trans*-retinol is rapidly released by photoreceptor cells to the IPM, where the interphotoreceptor retinoid-binding protein (IRBP) secreted by photoreceptors in this extracellular space facilitates binding to the RPE (reviewed in Travis et al, 2007). Here, the major retinyl ester synthase in RPE cells, lecithin-retinol acyl transferase (LRAT), esterifies all-*trans*-retinol to a fatty acid and the retinyl ester is isomerised and de-esterified by a single protein in the RPE, known as RPE-specific 65-kDa protein (RPE65) (reviewed in Travis et al, 2007; Lee et al, 2010). This reaction generates 11-*cis* retinol, which is oxidised to form 11-*cis*-retinal, which is transported back to the photoreceptor outer segment to regenerate rhodopsin (reviewed in Lee et al, 2010). The regeneration of rhodopsin is crucial for its role in light activation in the phototransduction pathway, resulting in night vision (reviewed in Kijas et al, 2002) and in the renewal of photoreceptor outer segments (Young, 1967).

1.6 Pathways of photoreceptor degeneration

Photoreceptor cell development, physiology, maintenance, and the genes implicated in specific pathways contributing to these factors can provide further understanding of retinal degenerations. Transcriptome studies have revealed that over 10,000 genes (> 35% of the canine genome), are expressed in canine retinal tissue (Reddy et al, 2015). These genes are involved in a myriad of biological pathways and processes, all of which are necessary to ensure normal development and optimal performance of the retina. Thus, mutations in genes involved in vision related processes may cause a loss-of-function of these genes and therefore possibly a loss-of-function in the retina, resulting in retinal dysfunction. However, the complexity of the vertebrate retinal structure and function mean that many mechanisms by which mutations lead to photoreceptor degeneration and blindness are not completely understood. Broadly, four main categories can be defined as mechanisms of degeneration (Pierce, 2001):

a) Disruption of the photoreceptor outer segment
b) Metabolic overload
c) Defects in the RPE
d) Continuous activation of phototransduction

Further exploration to increase understanding of these mechanisms may help targeted gene therapies for degenerative retinal diseases.

Disruption of the photoreceptor outer segment

As previously described, photoreceptor outer segments are formed from the primary cilia, therefore mutations in genes encoding ciliary compartment proteins disrupt photoreceptor ciliary function in photoreceptor cells. Examples of this include mutations in the retinitis pigmentosa GTPase regulator-interacting protein 1 gene (*RPGRIP1*), associated with human Leber congenital amaurosis (LCA) and retinitis pigmentosa (RP) as well as canine cone-rod dystrophy-1 (cord1) (Mellersh et al, 2006). RPGRIP1 localises in the connecting cilium of photoreceptors, and is required for outer segment development, cilia protein trafficking and disc morphogenesis (Boylan & Wright, 2000; Zhao et al, 2003). This is exemplified by histological examination of the retinas of Miniature Long-Haired Dachshund dogs with cord1, of which findings included thinning of the outer nuclear layer, shortening of the photoreceptor outer segments and defective rod outer segment disc development from as early as 10.5 weeks of age (Curtis & Barnett, 1993). Another example of abnormal retinal development led by the disruption of photoreceptor outer segment morphogenesis is caused by mutations in the rhodopsin gene (*RHO*) in humans (Dryja et al, 1990; Iakhine, 2004; McWilliam et al, 1989) and other species. An accumulation of extracellular vesicles around photoreceptors, probably a product of aborted outer segment formation, and attenuation of outer segments were observed in RHO-null mice (Li et al, 1996), with some mice showing no outer segment formation at all (Lem et al, 1999). Similar findings have been demonstrated in RHO-null pigs (Ross et al, 2012) and dogs. Electroretinogram (ERG) evaluations of English Mastiff dogs harbouring a mutation in *RHO* exhibited clinical signs similar to that of human autosomal dominant RP, with attenuation of rod photoreceptor outer segments (Kijas et al, 2002). Furthermore, knockout of the gene encoding tubby-like protein 1 (*TULP1*) in mice exhibited progressive retinal degeneration with a unique feature of accumulation of extracellular vesicles in the IPM of the RPE; abundant toward the proximal inner segments and outer segments of photoreceptors (Hagstrom et al, 1999). A single defect in complex pathways involved in photoreceptor outer segment morphogenesis could affect protein-trafficking regulation resulting in these retinal phenotypes described.

Metabolic overload

Metabolic overload describes an accumulation of ions within a cell leading to the maintenance of a constant electrochemical gradient, which becomes toxic to the cell (Pierce, 2001). In photoreceptor cells, this is probably due to a continuous high demand of the cGMP-gated channels where a constant influx of Na^+ and Ca^{2+} is directed across the outer segment membrane. As observed in canine and murine models, mutations in the gene encoding rod cGMP phosphodiesterase beta subunit (*PDE6B*) lead to excessive cGMP levels, resulting in a loss of photoreceptor cells. In Irish Setter dogs with a canine retinal disease, rod-cone dysplasia type 1 (RCD1), photoreceptor loss is detected as early as postnatal day 10 (Clements et al, 1993; Farber & Lolley, 1974; Pittler & Baehr, 1991). High cGMP levels promote the cGMP-gated channels to

remain open, allowing an influx of Na^+ and Ca^{2+} through the phototransduction cascade and cause metabolic overload in the outer segment (Pierce, 2001; Travis, 1998). Important mutations associated with high cGMP levels leading to photoreceptor cell death include genes encoding cyclic nucleotide gated (CNG) channels in the outer segment membrane, demonstrating a link between Ca^{2+} influx and rod photoreceptor apoptosis in mutant mice (Paquet-Durand et al, 2011). In addition, a mutation in the gene encoding human membrane guanylyl cyclase 1 (*GUCY2D*) in transgenic mice showed increased cGMP binding to CNG channels and therefore depolarisation of the outer segment due to an influx of Na^+ and Ca^{2+}. Eventual accumulation of Ca^{2+} concentration in mice rod outer segments triggers photoreceptor cell death (Sato et al, 2018).

Defects in the RPE

Photoreceptor outer segments shed on a regular basis and undergo phagocytosis by adjacent RPE cells, recycling metabolites to photoreceptors. Defects in genes encoding proteins that are implicated in the phagocytosis of shed disc tips by RPE cells result in retinal dysfunction in rodents, humans and dogs. One example includes the Mer tyrosine kinase proto-oncogene, *MERTK,* whereby mutations in *MERTK* cause RP through defective outer segment phagocytosis by the RPE (D'Cruz et al, 2000; Gal et al, 2000; Everson et al, 2017). This was first demonstrated in a rat model of autosomal recessive RP where histological examination of the rat retina displayed RPE dysfunction and an abnormal accumulation of photoreceptor outer segment debris between the photoreceptors and RPE, resulting in photoreceptor cell death (Dowling & Sidman, 1962; Vollrath et al, 2001). Another example of genetic mutations that implicate pathways within the RPE include *RPE65*, where RPE65 is abundant in the RPE and converts all-trans retinyl ester to 11-cis retinol (Moiseyev et al, 2005), which is essential for normal retinal function. Mutations in *RPE65* are associated with inherited retinal degenerations in humans (Bowne et al, 2011; Gu et al, 1997; Marlhens et al, 1997) and dogs (Aguirre et al, 1998; Bechynova et al, 2008; Veske et al, 1999), resulting in defective RPE cells and photoreceptor apoptosis.

Continuous activation of phototransduction

As previously discussed, the phototransduction cascade ensures proper cycling of ions to maintain electrochemical homeostasis in the retina. Disruption of proteins triggering these pathways and channels can alter the cascade of biochemical reactions, leading to photoreceptor apoptosis. Continuous activation of the phototransduction cascade alters the energy utilisation of the retina, triggers or inhibits enzyme activity, limits cation entry into photoreceptor outer segments and alters the concentration of signalling molecules within the cascade, in particular Ca^{2+} (Fain, 2006). Mutations in genes encoding components of the machinery in the phototransduction cascade result in continuous activation of this process and lead to retinal degeneration in humans and other species (Chen et al, 1999; Sullivan et al, 2017). Once light activates rhodopsin in the phototransduction cascade it must be turned off by two significant

proteins: rhodopsin kinase and arrestin. Knockout mouse models of rhodopsin kinase (Chen et al, 1999) and arrestin (Xu et al, 1997) demonstrates how rods lacking these proteins show altered rod light responses and light-induced apoptosis. Both models demonstrate slow retinal degeneration that was accelerated by chronic activation of phototransduction when mice are exposed to a constant light stimulus. Thus, the arrestin gene, also known as S-antigen visual arrestin-1 (*SAG*), functions in inactivating rhodopsin during the phototransduction cascade. Although a mutation in this gene is associated with a late-onset hereditary retinal degeneration in Basenji dogs, further study into canine *SAG*-mutations is required to determine if similar mechanisms are affected, as demonstrated in knockout mouse models.

1.7 Human inherited retinal degenerations (IRDs)

Human inherited retinal degenerations (IRDs) encompass a diverse group of progressive disorders characterised by loss of rod and/or cone photoreceptors eventually leading to visual impairment and blindness. They are clinically and genetically heterogeneous, where mutations in genes critical to retinal function are responsible for disrupting pathways and mechanisms leading to photoreceptor cell death. IRDs can be clinically categorised based on the rate of progression and the types of retinal cells that are primarily involved in disease pathogenesis. The wide heterogeneity of human IRDs is illustrated by the large number of genetic mutations and number of genes associated with IRDs. In addition, there is extensive variation in clinical expression even among individuals sharing identical mutations. Despite the identification of 271 genes associated with IRDs (RetNet, 2020), many patients still lack a molecular diagnosis indicating that further elucidation of genes involved in these diseases is required.

Retinitis pigmentosa (RP)

The first form of retinal degeneration in humans was described in 1857 as night blindness, termed as retinitis pigmentosa (RP) (Donders, 1857). Today, RP is the most common form of IRD, affecting 1 in 4,000 people worldwide before middle age (Verbakel et al, 2018). The disease first manifests as reduced vision in dim light (nyctalopia) caused by primary depletion in rod photoreceptors. As the disease progresses, cone photoreceptors degenerate leading to a gradual reduction in vision and total blindness. Typical fundus abnormalities include attenuation of retinal blood vessels, a waxy pallor of the optic nerve and mottling and granularity of the RPE. Intraretinal bone-spicule pigmentation presents due to pigment migration to and accumulation in the retina, due to breakdown of RPE cells. A definitive diagnosis requires an ERG to reveal loss of photoreceptor function and to determine the severity in early stages of the disease. RP patients exhibit narrowing of the visual field in the dark, therefore typically have reduced vision in dim light or in the dark and have a progressive loss of their peripheral vision. Central visual function diminishes throughout disease progression. Colour vision generally remains good until visual acuity worsens; however, this can deplete if cone dysfunction occurs early in the disease (reviewed in Weleber &

Gregory-Evans, 2006). To date, 63 genes have been associated with autosomal recessive RP; 30 genes associated with autosomal dominant RP; and three genes and loci associated with X-linked RP (Appendix 1) (RetNet, 2020). In general, RP-associated genes are known or expected to be expressed in retinal photoreceptor cells or the RPE and impact the proper function of the visual cycle.

Leber congenital amaurosis (LCA)

The most severe human IRD is LCA, primarily affecting both rod and cone photoreceptors accompanied by changes in the RPE within the retina. LCA is described as an early-infantile onset rod-cone dystrophy, presenting as congenital or within the first few months of life. Early-onset severe retinal dystrophy (EOSRD) is considered a milder form of LCA, with clinical signs becoming apparent after infancy and usually before the age of five years (Kumaran et al, 2017). Clinical signs of LCA include nystagmus (repetitive, uncontrolled eye movement), poor pupillary light response and an abnormal fundus appearance, similar to RP. An LCA diagnosis is apparent from undetectable ERG responses due to a lack of photoreceptor response whereas EOSRD show photoreceptors are less affected with minor ERG signals (Kumaran et al, 2017). There is overlap between genes involved in LCA and EOSRD, however, mutations in specific genes are more often associated with one disease or the other. To date, three genes have been identified to cause autosomal dominant LCA and 23 genes associated with autosomal recessive LCA (Appendix 1) (RetNet, 2020).

Ciliopathies

Ciliopathies encompass a group of diseases caused by mutations in genes involved in ciliary function, leading to ciliary defects. Cilia are a nearly ubiquitous component of all vertebrate cells therefore ciliary defects can affect many organ systems. Many ciliary proteins are involved in crucial roles in animal physiology, signalling and development, thus cilia-related disorders manifest as a wide spectrum of features including retinal degeneration, kidney abnormalities, obesity and skeletal dysplasia (Fliegauf et al, 2007; Waters & Beales, 2011). The targeting and assembly of ciliary cargo is controlled by specialised multiprotein complexes within the cilium, including IFT proteins and, briefly above-mentioned, the BBSome. The BBSome is important in regulating protein trafficking in the primary cilium and plays a crucial role in photoreceptor outer segment morphogenesis and maintenance (Hsu et al, 2017). Mutations in genes encoding IFT or BBS proteins, as well as other cilium associated genes, result in ciliopathies. Specifically, mutations in BBSome proteins cause human Bardet-Biedl syndrome (BBS), a ciliopathy where patients present with retinal degeneration and additional features including diabetes, obesity, polydactyly, renal dysfunction and dental crowding. Currently, BBS-associated variants have been identified in 25 genes (Appendix 1) (RetNet, 2020). Other syndromic forms of RP associated with

ciliopathies include Usher syndrome, Cohen syndrome, Joubert syndrome, Senior Loken syndrome and Sensenbrenner syndrome (Verbakel et al, 2018).

There are many proteins that localise to the photoreceptor ciliary compartment, e.g. FAM161A, NPHP5 and RPGRIP1, and mutations in genes encoding these proteins have been implicated in the canine IRD: progressive retinal atrophy (PRA) (Downs & Mellersh, 2014; Goldstein et al, 2013b; Mellersh et al, 2006). PRA will be described further below. BBS genes have also been associated with canine PRA (Chew et al, 2017; Downs et al, 2014). Studies in humans and animal models have shown that ciliary genes are essential for photoreceptor outer segment development, homeostasis and photoreceptor survival (Downs et al, 2016; Karlstetter et al, 2014; Won et al, 2009).

1.8 Canine inherited retinal degenerations (IRDs)

Canine IRDs present similarly to human forms and represent a group of inherited retinal diseases associated with reduced photoreceptor function leading to visual impairment and eventually blindness. The first report of a canine IRD by Magnusson (1911) identified it as the canine homolog to human RP, sharing similar clinical and genetic features and rates of progression. Classifying canine IRDs is complex, however they are typically categorised depending on which photoreceptor cells deplete initially, with further classification based on the rate of progression and when initial signs of disease present (early- and late-onset). Canine IRDs can be characterised into two broad clinical categories: stationary or progressive.

Stationary IRD

The first stationary retinal dystrophy to be characterised was congenital stationary night blindness (CSNB) in Briard dogs and had similarities to a human CSNB (Narfström et al, 1989). Affected dogs displayed congenital night blindness, with normal fundi up to 2-3 years of age and reduced ERG responses. Since this initial report, the disease has been further defined and renamed as 'retinal dystrophy' as progressive changes were apparent (Narfström, 1999). A 4-bp deletion in the *RPE65* gene has been identified in purebred Briard dogs and Briard crossbreeds presenting with CSNB/retinal dystrophy (Aguirre et al, 1998; Veske et al, 1999). However some Briard dogs displaying retinal abnormalities consistent with PRA have been identified that do not carry the *RPE65* mutation, and have no other provocative variants in this region (Aguirre et al, 1998), illustrating that this breed segregates two genetically distinct forms of retinal degeneration.

Cone degeneration is phenotypically similar to a human disease, achromatopsia: a heterogeneous autosomal recessive disorder (Dixon, 2016; Seddon et al, 2006). A cone degeneration described in Alaskan Malamute dogs specifically affects cone photoreceptors leading to day blindness (hemeralopia). Clinically, cone-degenerate puppies develop hemeralopia and light sensitivity (photophobia) between 8-12 weeks of age, with normal vision in dim light. ERG readings detect loss of cone function from three weeks of age and as affected dogs mature, rod function remains

normal. Affected individuals homozygous for mutations in the gene encoding cyclic nucleotide gated channel β-3 (*CNGB3*) have been identified in Alaskan Malamutes and German Shorthaired Pointer dogs (Sidjanin et al, 2002). The deletion removing all exons of canine *CNGB3* reported initially in Alaskan Malamutes was also detected in miniature Australian Shepherd dogs (homozygous); and one Siberian Husky and three Alaskan sled dogs were identified as carriers (Yeh et al, 2013). Further to this, a mutation in the cyclic nucleotide gated channel α-3 gene (*CNGA3*) was identified in German Shepherd dogs with cone degeneration, presenting the dog as a canine model to aid in the identification of human achromatopsia-associated channelopathies and treatments (Tanaka et al, 2015).

Progressive Retinal Atrophy (PRA)

Canine PRA is the term given to a group of heterogeneous diseases characterised by depletion in photoreceptor cells in the retina over time. Progression of the disease is variable, as well as the aetiology and age of onset, which can be broadly divided into early- and late-onset. PRA shows phenotypic similarities to that of human RP. Clinical signs revealed by ophthalmoscopic examination include bilateral vascular attenuation of retinal blood vessels, becoming more marked as the severity of the degeneration progresses, and retinal thinning leading to hyperreflectivity of the tapetum lucidum. In moderate to advanced stages of disease, depigmentation and mottling of the non-tapetal fundus can also be observed. The optic disc typically turns pale in colour due to loss of retinal circulation. Typically, nyctalopia is exhibited and in many cases bilateral secondary cataracts develop (Parry, 1953; Petersen-Jones, 2005). Ophthalmoscopic changes may be supported by ERG readings, measuring the rod and cone photoreceptor response. Sophisticated ERG protocols allow the differentiation of progressive rod-cone degenerations (RCDs) from cone-rod degenerations (CRDs), which can prove difficult using ophthalmoscopy alone (Acland, 1988). Where a distinction can be made between RCDs and CRDs, RCDs are defined by initial loss of rod photoreceptors responses followed by that of cones. As rod photoreceptor cells function in low light conditions, with low spatial acuity, impaired vision in low level light and darkness manifest initially in rod led degenerations. Responsible for colour vision, cones cells are active at higher levels of light and degeneration of these cells follows rod depletion, causing visual impairment in bright light (reviewed in Gelatt, 2013). Forms of RCDs previously reported display similar clinical signs, although variability in rates of progression, age of onset and aetiology are exhibited across and within breeds. This is shown across previous studies in breeds including the Miniature Poodle, Norwegian Elkhound, Briard, Miniature Schnauzer, Gordon and Irish Setter, Cardigan Welsh Corgi, Sloughi and Rough and Smooth Collie (Acland & Aguirre, 1987; Aguirre et al, 1982; Clements et al, 1993; Dekomien et al, 2000; Downs et al, 2013; Kukekova et al, 2009; Millichamp et al, 1988; Parshall et al, 1991; Petersen-Jones et al, 1999). In CRDs cones are predominantly affected followed by the depletion of rods. Although clinical differences occur, CRDs demonstrate ophthalmoscopic changes that are very similar to those of

RCDs. Fewer breeds are affected with CRDs, yet those that are show severe retinal and fundus abnormalities and ultimate vision loss (Curtis & Barnett, 1993; Kijas et al, 2004; Ropstad et al, 2007a; Ropstad et al, 2007b). ERG assessment is not routinely performed in dogs, so distinguishing between a rod-cone or cone-rod degeneration is difficult. However, when night blindness is the first clinical sign observed, a rod-cone degeneration is considered the most likely diagnosis. The lack of ERG assessment in dogs commonly leaves ophthalmoscopic examination as the extent of the diagnosis, and has resulted in the use of the umbrella term 'PRA' to describe both RCDs and CRDs in dogs.

Early-onset PRA-affected individuals first exhibit clinical signs between six weeks and six months of age, typically leading to blindness between the ages of 1-5 years, thus sharing clinical similarities with human LCA. Genetically distinct, early-onset forms of RCDs include rod-cone dysplasia type-1 (RCD1) (Clements et al, 1993; Dekomien et al, 2000), type-2 (RCD2) (Kukekova et al, 2009), type 3 (RCD3) (Petersen-Jones et al, 1999) and early retinal degeneration (ERD) (Goldstein et al, 2010). The first canine PRA-associated mutation to be identified was associated with a form of RCD in the Irish Setter dog: RCD1 (previously described). Initial presentation of RCD1 in Irish Setters demonstrated reduced ERG measurements, with rod responses increasingly diminished in the early stages of retinal development and the degree of cone response less so. Morphologically, visual cells from the retinas of affected dogs are abnormal with a reduction of rods in the photoreceptor layer, short inner segments and fewer visual cells. Cones are affected to a lesser extent, still prominent with broad inner segments but abnormal, short outer segments. The inner retinal layers appear to remain normal, yet all rod cells degenerate by five months of age (Aguirre, 1978). A nonsense mutation in the *PDE6B* gene was identified as a causative mutation of autosomal recessive RCD1, resulting in a loss of function of rod PDE (Clements et al, 1993).

A novel PRA in Miniature Long-Haired Dachshund dogs was first clinically characterised as a CRD, termed cord1 (Curtis & Barnett, 1993). Histopathology showed the retinas of affected puppies were normal at six weeks of age, with retinal abnormalities detected from as early as 10.5 weeks of age. This suggested that retinas developed normally, and pathological changes developed after retinal maturation, indicative of a retinal degeneration rather than a retinal dysplasia. ERG assessment revealed clinical signs of PRA were apparent from as early as six months of age, suggesting an early-onset PRA (Curtis & Barnett, 1993). The genetic cause of this form of PRA in the Miniature Long-Haired Dachshund was later identified as a 44-bp insertion in exon 2 of the *RPGRIP1* gene (Mellersh et al, 2006). Despite complete correlation of the *RPGRIP1* mutation with clinical disease in the initial Miniature Long-Haired Dachshund research colony, additional studies in a wider population discovered phenotype-genotype discordance, where vision defects and retinal degeneration in dogs developed at varied ages, or not at all within their lifespan (Miyadera

et al, 2009), more alike to *RPGRIP1* associated human retinopathies. Further ophthalmic studies into Miniature Long-Haired Dachshunds homozygous for the *RPGRIP1* insertion revealed some dogs developed clinical signs later in life, contradicting the previously reported findings of strictly an early-onset PRA and demonstrating that the *RPGRIP1* mutation alone does not invariably lead to an early-onset phenotype (Busse et al, 2011). An age of onset modifier locus was identified on canine chromosome 15 (CANFA15) in Miniature Long-Haired Dachshund dogs homozygous for the *RPGRIP1* mutation, where visual defects presented before the age of four years (Miyadera et al, 2012). Further investigation later revealed a mutation in the microtubule-associated protein 9 (*MAP9*) gene as the age of onset modifier mutation inherited with the *RPGRIP1* insertion in Miniature Long-Haired Dachshund dogs with early-onset CRD (Forman et al, 2016).

Late-onset PRAs are characterised by initial normal retinal development, with photoreceptor degeneration occurring from around the age of five years. Dogs inheriting late-onset forms of disease show normal ERG responses at a young age, with progressive rod depletion later in their lives. As with early-onset RCDs, cones are abnormal and a reduction in cone response follows that of rods, although cone degeneration is less severe. Late-onset forms of PRA tend to involve genes that are significant for the maintenance and function of photoreceptors throughout life as opposed to early-onset forms for which genes necessary for postnatal development are implicated (Mellersh, 2014). Perhaps the most common late-onset PRA in dogs is progressive rod-cone degeneration (PRCD) a recessively inherited PRA first reported in Miniature Poodles (Aguirre et al, 1982) where the disease phenotype was characterised by aberrant photoreceptor structure. A single nucleotide substitution in the novel gene *PRCD* was identified as the causal mutation for PRCD initially across 18 breeds (Zangerl et al, 2006). At present, a commercially available DNA test is available for 72 dog breeds (Mars Veterinary, UK) with the mutation proving to be shared across many diverse breeds and is generally the first PRA-variant excluded when a distinct form of PRA is identified in a breed. This suggests that a common ancestral chromosome containing the mutant gene was widespread through the development of modern dog breeds. Following variant identification in the dog, a human RP patient was identified that was homozygous for the same mutation, establishing *PRCD* as a novel retinal gene causing retinal degeneration in dogs and humans (Zangerl et al, 2006).

In clinical and genetic studies, it is also important to consider the presence of PRA phenocopies: eye conditions that are not hereditary but share similar clinical signs to PRA. One example is sudden acquired retinal degeneration (SARD): a non-inherited retinal degeneration, where ophthalmoscopic signs of retinal degeneration are absent at the onset of blindness and appear many weeks following vision loss (Petersen-Jones, 2005). PRA phenocopies can complicate a PRA diagnosis and possibly lead to a misdiagnosis. Genetic investigations are reliant on a robust clinical diagnosis to ensure the correct disease status is applied to individuals being used for

research i.e. cases (affected dogs) and controls (non-affected dogs) for case-control studies. For the majority of PRA studies, all dogs selected to contribute to a study have been subject to an eye examination by a certified veterinary ophthalmologist, a panellist of the British Veterinary Association/Kennel Club/International Sheep Dog Society (BVA/KC/ISDS) Eye Scheme or the European equivalent.

1.9 Identification of disease-associated mutations

Single-gene diseases, often referred to as Mendelian, are usually inherited in one of several patterns depending on the location of the gene and whether one or two normal copies of the gene are required for the disease phenotype to manifest. There are five basic modes of inheritance for single-gene diseases: autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive, and mitochondrial. In autosomal dominant diseases, each affected individual has one copy of the disease-associated mutation inherited from an affected parent. Individuals carrying two copies of the disease-causing mutation will be affected with an autosomal recessive disease, presenting clinical signs of the disease; both parents are heterozygous (carriers) carrying one copy of the mutated gene and are clinically non-affected. In X-linked diseases, the X chromosome harbours the causal mutation. Females are more frequently affected in X-linked dominant inheritance and, although affected males and females can occur in the same generation, fathers cannot pass the trait to male progeny. X-linked recessive diseases present more frequently in males, often in each generation. Both parents must carry one affected X chromosome to produce affected female offspring, where only a carrier female parent can pass the disease to male offspring (a male parent cannot pass an X-linked trait to its male offspring). Mitochondrial diseases can affect both males and females, however only female individuals will pass on the disease to offspring through maternal inheritance (reviewed in Genetic Alliance, 2010).

As with many canine Mendelian disorders, the majority of canine PRAs reported are autosomal recessive, however, one dominant (Kijas et al, 2002) and four X-linked forms (Kaukonen et al, 2020; Kropatsch et al, 2016; Vilboux et al, 2008; Zhang et al, 2002) have also been described. Canine PRAs are genetically heterogeneous, where multiple mutations exist in individual breeds that are deemed disease-causing, though these may not account for all PRA cases in the affected breeds. To date, 27 mutations in over 60 breeds are associated with PRA (Appendix 2); with many more yet to be elucidated. Although there are PRA-associated mutations that have been detected across multiple diverse breeds, e.g. PRCD (Zangerl et al, 2006) identified in at least 65 breeds to date and rod-cone degeneration 4 (RCD4) identified in Gordon Setter, Irish Setter, English Setter, Tibetan Terrier and Standard Poodle dogs (Downs et al, 2013), the majority of reported canine PRAs are often private to individual breeds. Knowledge of the mode of inheritance of a single-gene disorder drives the choice of genetic application(s) used in a study. A variety of genetic techniques

have been used to identify previously reported causative PRA-variants; some of which will be further discussed.

Linkage analysis

The fundamental basis of linkage analysis is a family-based analysis, requiring DNA from the proband and an informative extended family pedigree with other affected family members. This is a genetic mapping approach used to estimate the distance between loci or genes at different locations on a chromosome. Linkage analysis was predominantly used prior to high density SNP arrays and next-generation sequencing (NGS) technologies to map disease-associated regions. Suitable polymorphic markers used in a linkage analysis approach include polymorphic microsatellites and SNPs. Microsatellites are short repetitive DNA sequences that vary in length and are found interspersed throughout the genome. These range in length from one to six or more base pairs that are repeated, however typically consist of di-, tri- and tetra- nucleotide repeats. Dinucleotide repeats e.g. $(CA)_n$ occur most frequently and are more stable in the canine genome than trinucleotide or tetranucleotide repeats (Francisco et al, 1996). These consist of at least 12 dinucleotide repeats, and longer repeats are associated with greater allelic variation, due to the presence of multiple length alleles: the longer the stretch of repeats, the more polymorphic the repeat is. The high variability of microsatellites makes them desirable in linkage analysis studies as they are highly polymorphic. Another tool used in a linkage analyses are SNP arrays. SNP arrays enable genotyping of thousands of polymorphic SNP markers spanning the genome to identify those that segregate with a disease in family-based pedigrees to identify disease-associated loci and a disease-associated haplotype. During meiosis, the process of recombination occurs during which homologous chromosomes align, break to exchange genetic material and re-join in new combinations of alleles, different to both parents. Recombination is more likely to occur between two markers positioned far apart on different chromosomes, which are independent of one another, than between two markers close to each other on the same chromosome. The basis of linkage analysis assumes that alleles of markers physically located close together are inherited together, and when inherited with a disease can identify the disease-associated locus. Successful use of these approaches is demonstrated in a study of Miniature Long-Haired Dachshund dogs with a CRD, where a microsatellite-based homozygosity mapping followed by linkage analysis was used to identify a disease-associated locus on CANFA15. Further interrogation of the locus revealed a 44-bp insertion in exon 2 of *RPGRIP1* (Mellersh et al, 2006).

Candidate gene study

Knowledge of genes implicated in previous studies of a disease can be adopted for a candidate gene study where there is previous evidence that a gene is associated with a disease of interest, often in another breed or species. This can be used in conjunction with a linkage approach, where selecting polymorphic markers that are located as close as possible to a gene of interest will

determine if both are in LD and are within a disease-associated region. In some instances, candidate genes are canine homologues of those implicated in similar human diseases, but genes can also be selected if they fall within chromosomal regions that are homologous between species. For example, canine X-linked PRA (XLPRA) is akin to a human X-linked RP (RP3) and maps to the same chromosomal location in the X chromosome (Zeiss et al, 2000; Zhang et al, 2001). A study by Zhang et al. (2002) employed a linkage analysis approach to map two canine XLPRAs to the region and identified mutations associated with XLPRA type-1 (XLPRA1) in the Siberian Husky and Samoyed dogs, and XLPRA type-2 (XLPRA2) in mongrel-derived dogs. The previously discussed *PRCD* mutation is another example of how canine research has informed human medicine in retinal degeneration. After the mutation in *PRCD* was initially detected in dogs the gene was screened in unresolved human RP cases, and a patient with the identical mutation was subsequently identified (Zangerl et al, 2006). RetNet, the Retinal Information Network (RetNet, 2020), is an online database summarising all genes and loci identified in human retinal diseases, including RP. Thus, many of these genes are candidate genes for canine PRA studies, or phenotypically similar diseases. Another reason for a gene to be considered as a candidate for a disease is the known function or involvement with relevant biological mechanisms and pathways of the encoding protein (examples of such pathways are discussed in section 1.6). This can highlight novel candidate genes implicated in disease. Examples of candidate gene studies include the identification of a 1-bp deletion in *PDE6A* in PRA-affected Cardigan Welsh Corgi dogs (Petersen-Jones et al, 1999) and an 8-bp deletion in *PDE6B* in PRA-affected Sloughi dogs (Dekomien et al, 2000). Candidate gene approaches were the method of choice prior to the availability of genome-wide methods, such as genome-wide association studies.

Genome-wide association study (GWAS)

A genome-wide association study (GWAS) employs a large-scale SNP array that genotypes hundreds of thousands of SNPs across a genome, and tests these SNPs for statistical associations with a trait or disease. The most common approach in the dog utilises a case-control study, where the frequency of both alleles at biallelic SNP marker loci is compared between affected individuals (cases) and non-affected individuals (controls), and statistical methods are applied to determine the likelihood that an observed difference has occurred by chance. A statistically significant association between a given disease phenotype and SNP markers suggests that the markers are associated with the disease. This can narrow a disease associated region from the entire genome down to a small region, in the dog usually a few Mb in size for Mendelian conditions because of the aforementioned extensive LD in dog breeds. Various SNP arrays are commercially available to genotype SNP markers in many species. For the dog, popular SNP arrays include two Illumina CanineHD Whole-Genome Genotyping BeadChips: the 170K chip containing 172,115 highly polymorphic SNP markers, based on the earlier CanFam2.0 reference sequence; and the 230K chip containing approximately 230,000 SNPs, which is an updated version on the 170K chip. Both these

arrays have been successfully used in studies of monogenic diseases, but less so for polygenic/complex conditions. More recently, a higher density array has become available: the Axiom™ Canine Genotyping Array Set (Thermo Fisher Scientific, Loughborough, UK) includes more than 460,000 markers. The increased number of polymorphic markers on this array makes it more suitable than the two Illumina arrays for investigating inherited complex conditions where multiple more common/ancestral risk factors may be contributing to a disease phenotype. SNPs on these dense arrays were selected as they are polymorphic across a variety of dog breeds to ensure their utility across breeds. As previously discussed, dog breeds show high levels of LD (Sutter et al, 2004) therefore the use of GWAS mapping approaches has proven successful in identifying candidate regions and subsequently mutations for hereditary retinal diseases (Goldstein et al, 2006). Moreover, the presence of LD blocks and low intrabreed diversity means fewer SNP markers and smaller sample sizes can be used in canine GWAS than would be required to perform the same experiment in human populations. This has successfully been demonstrated in PRA studies using modest numbers of cases and controls (e.g. between 6-27 cases and 3-29 controls) with a GWAS approach to identify causative mutations (Ahonen et al, 2013; Downs et al, 2011; Downs et al, 2013; Downs et al, 2014; Downs & Mellersh, 2014; Goldstein et al, 2013a; Kropatsch et al, 2010; Murgiano et al, 2018; Wiik et al, 2008; Wiik et al, 2015). Although fewer SNP markers might be necessary to identify a region of LD surrounding a disease locus, the region of LD will be longer than the equivalent region in the human genome, therefore the fine mapping of a disease locus is often more difficult.

Next-Generation Sequencing

While Sanger sequencing is considered the gold standard for the identification of disease-associated variants, this is a laborious and costly method when exploring numerous genes with many exons, or an entire genome. High throughput methods have been developed relatively recently to sequence more DNA at a reduced cost and faster rate. Next-generation sequencing (NGS), also referred to as massively parallel sequencing, is a high-throughput approach to DNA sequencing. The emergence of NGS has revolutionised genetic studies, including the identification of disease-associated variants, across many species. Two types of NGS applications are often used in the study of inherited disease: whole exome sequencing (WES) and whole genome sequencing (WGS).

Whole exome sequencing (WES)

Whole exome sequencing (WES) targets the protein-coding regions of genomic DNA (gDNA), known as the exome. The exome represents an enriched subset of the genome accounting for approximately 2% of the canine genome (Broeckx et al, 2014). A WES approach allows an inexpensive means of identifying functional variants associated with disease and is based on the use of baits to enrich and target specific exonic regions to sequence. As many Mendelian diseases

previously reported are caused by variants that disrupt protein-coding or splice site regions, several studies have employed WES to identify genes associated with inherited retinal disease, when prior mapping information is unknown (Xu et al, 2015; Zuchner et al, 2011). In human IRD studies WES is currently a popular method of choice to identify novel genes when a candidate gene approach does not yield causative variants (Siemiatkowska et al, 2014). However, WES does not detect intronic variation, including structural variants such as inversions where breakpoints may be within intronic regions; and deep intronic variants, which may cause aberrant splicing events, create cryptic splice sites, or disrupt regulatory elements (Bax et al, 2015; Webb et al, 2012). For such variants, targeted genomic sequencing or whole genome sequencing (WGS) approaches may be more appropriate.

Whole genome sequencing (WGS)

During the last decade, WGS is increasingly becoming a popular method in disease variant identification, either as a stand-alone method or in combination with other technologies such as linkage analysis or GWAS. This is largely due to a decrease in costs over the past five years that has enabled multiple genomes to be sequenced for one study and comparisons carried out between cases and controls. By sequencing an entire genome, a more comprehensive data set is obtained compared to WES where only the coding regions are sequenced. In principle, WGS detects all variation, including structural variants and non-coding variants. The caveat of WGS is that no positional information can be determined regarding where a causal mutation will lie, whereas a GWAS can highlight chromosome(s) harbouring the mutation to further define a disease associated region. Despite this, if finances enabled WGS of many cases and controls, WGS data could be used for a GWAS by extracting GWAS SNP positions. However, the cost per sample of WGS an individual is at least 10 times more than genotyping an individual for a GWAS using the aforementioned SNP arrays. Other disadvantages of WGS are the extensive data processing time and storage of a large amount of data. In addition, WGS analysis can be susceptible to bias if used in conjunction with a candidate gene approach; and non-exonic variants can be difficult to identify. With WGS analysis, various filtering steps must be undertaken to strategically decide which variants should be under consideration for involvement with the disease to reduce the initial high number of variants obtained.

1.10 The dog as a model for human IRDs

The study of canine retinal disease offers opportunities for novel target gene discovery in human RP and other retinal degenerations, where a large proportion of patients still have an unknown molecular diagnosis. Similarities between the human and canine eye morphology and size means that many surgical approaches and treatments can be trialled and performed in canine and human eyes. In addition, as both species reside in similar environments, the dog eye can be a useful model for human retinal research.

Naturally occurring canine models exist for several IRDs, some of which have contributed to gene augmentation therapies in humans such as the *RPE65* model for LCA. *RPE65* is expressed in the RPE and is a crucial component in the visual cycle, which regenerates visual pigments for phototransduction. The retinas of Briard dogs homozygous for a 4-bp deletion in *RPE65* show an accumulation of lipoidal inclusions within the RPE, with abnormal ERG responses suggesting a defect in the phototransduction pathway (Aguirre et al, 1998). This results in a slowly progressive photoreceptor degeneration sharing profound visual defects to human LCA. The introduction of adeno-associated virus (AAV) vectors carrying the normal *RPE65* cDNA enables gene transfer to *RPE65* mutant dog retinas and demonstrates rescued RPE function and overall retinal function, specifically in the visual cycle (Acland et al, 2001). In Briard dogs under observation, ERG responses were improved with the therapy providing evidence of long-term improved function of photoreceptors (Acland et al, 2005). Following success of canine *RPE65* gene therapy studies, human clinical trials have shown the same benefits in *RPE65* gene therapy using AAV (Bainbridge et al, 2008; Russell et al, 2017; Simonelli et al, 2010). This has led to the US Food and Drug Administration approving the treatment of patients with *RPE65*-mediated IRD using this method of gene therapy since January 2018. In 2019, this treatment was also recommended for use in the NHS in the UK, with the first patient receiving treatment in February 2020 at Moorfields Eye Hospital in London.

Gene therapy for canine models of RP include studies into the *PDE6A* (Mowat et al, 2017) and *PDE6B* (Petit et al, 2012) autosomal recessive RP phenotypes, and *RPGR* for X-linked RP (Beltran et al, 2012). As with the *RPE65* canine model, pre-clinical trials using AAV-mediated gene replacement therapy to inject vectors containing normal cDNA into mutant human retinas have shown to preserve and rescue photoreceptor function. Further clinical trials in humans are underway (Cehajic Kapetanovic et al, 2019). Canine models for human IRDs have provided proof of concept for treating human IRD, offering optimistic outlooks for patients with these blinding, incurable diseases.

1.11 Canine DNA testing

PRA is a canine welfare concern as vision loss is inevitable. Currently there are no treatments generally available for PRA in dogs, although, as discussed above, research studies using dogs have shown the effectiveness of the use of gene therapy as a treatment for some forms of retinal degeneration (Beltran et al, 2012; Lheriteau et al, 2014; Mowat et al, 2017; Occelli et al, 2017; Petit et al, 2012). In addition to gene augmentation therapies, cell-replacement therapy for human RP patients is being explored using the RNA-guided Cas9 nuclease from the microbial clustered regularly interspersed short palindromic repeats (CRISPR/Cas9) system to correct genetic mutations in human cell lines (Artero Castro et al, 2019), but these studies are in their relative infancy. The development of commercially available DNA tests for PRA-associated mutations

therefore plays an important role in controlling the incidence of PRA by enabling dog breeders to avoid breeding genetically affected dogs and thereby reduce the incidence of PRA in their breeds. Each DNA test is developed to target a specific genetic variant that has supporting evidence to cause disease in the breed of interest. For autosomal recessive conditions where two copies of the mutant allele are required for an individual to present with clinical signs of disease, genetic assays are developed to determine the genotype of each dog, identifying them as either: genetically affected (i.e. homozygous for two copies of the mutant allele), a carrier (i.e. heterozygous with one copy of the normal allele and one copy of the mutant allele), or genetically clear (i.e. homozygous for the wild type allele) of a disease-causing variant.

Clinical eye screening complements the use of DNA tests, where the former can identify novel or emerging eye conditions for which a genetic variant has not yet been discovered, and the latter enables dog owners and breeders to use a one-off genetic test to determine their dog's genotype with respect to a specific mutation. This knowledge facilitates informed breeding decisions and can detect subclinical cases prior to disease onset. This is especially important for diseases that may present as late-onset and past the typical breeding age of the dog. DNA tests can also identify non-affected heterozygotes (carriers) for a disease-associated mutation, which a clinical eye examination cannot. This is particularly important when considering a hereditary disease with a high allele frequency in a population, to ensure carrier dogs are bred appropriately (e.g. to genetically clear dogs) to prevent producing affected offspring. It is also important to preserve genetic diversity within a dog breed, which in some instances is achieved by keeping carriers in the gene pool. For some hereditary diseases, especially where the allele frequency is low, it is possible to eliminate the disease allele by eliminating carrier and affected dogs from the breeding population. This is not always achievable for numerically small breeds; therefore, breeding decisions should be made with caution and tailored to the breed in question. On the other hand, a mutation-specific DNA test does not screen for all disease-associated mutations identified in a breed. If other forms of ocular disease with unidentified mutations are present in the breed, DNA testing will not detect the disease, but clinical screening will.

1.12 The British Veterinary Association/Kennel Club/International Sheep Dog Society (BVA/KC/ISDS) Eye Scheme

The main purpose of the BVA/KC/ISDS Eye Scheme (https://www.bva.co.uk/Canine-Health-Schemes/Eye-scheme/) is to identify inherited and non-inherited eye diseases in dogs. Until January 2020, the scheme was comprised of two schedules: Schedule A containing breeds with confirmed inherited eye diseases and Schedule B containing breeds with eye diseases suspected as being inherited and under investigation, based upon the consistent opinion of veterinary ophthalmologists in the UK. Recent assessment has updated these to one schedule (Appendix 3)

where breeds listed are advised to undergo annual eye examinations carried out by a BVA/KC/ISDS panellists, with a result issued as 'clinically affected' or 'clinically unaffected'.

1.13 Thesis aims

In this thesis, novel forms of PRA will be explored in the Lhasa Apso (LA), Giant Schnauzer (GS) and Shetland Sheepdog (SS) dog breeds, independent of one another. Elucidation of novel PRA-associated variants aims to improve the understanding of the aetiology of distinct forms of PRA in these three breeds. The identification of genes responsible for such afflictions will provide further insight into the pathways and interactions involved in retinal function. In addition, the identification of PRA-associated variants will enable DNA tests to be developed based on any novel mutations identified to offer as a breeding tool, as PRA in many dog breeds continues to burden dog breeders. Moreover, identification of novel genes associated with canine PRAs will prompt their consideration as candidate genes for RP or other IRDs in humans and other species.

The major research objectives of this thesis are summarised as below:

1. Identify breeds affected with PRA for which the causal mutation is unknown.
2. Elucidate mutations associated with PRA in each breed under investigation, using whole genome investigations, including GWAS and WGS technologies, identifying the most efficient mutation discovery routes for differently structured proband groups.
3. Develop commercial DNA tests for all mutations identified.
4. Offer novel insights for consideration in human retinal degeneration research.

Chapter 2 General Materials and Methods

2.1 PRA diagnosis and sample collection

All PRA-affected dogs used in each study were examined by certified veterinary ophthalmologists through a clinical referral process or via the BVA/KC/ISDS Eye Scheme in the UK, or the European equivalent. Dogs with a PRA diagnosis were defined as "cases" and presented with clinical signs consistent with PRA. Dogs clinically clear of PRA were classed as "controls" and were deemed clear of inherited eye diseases upon ophthalmoscopic examination after the age of eight years, applying the same ophthalmoscopic examination methods as the cases, except where stated. Collection of DNA samples from animals using buccal mucosal swabs was approved by the Animal Health Trust (AHT) Ethics Committee (ref no. 24-2018E).

2.2 DNA extraction

DNA extraction from buccal mucosal swabs

To extract gDNA from buccal mucosal swabs, the QIAamp DNA Blood Mini or Midi Kits (Qiagen, Manchester, UK) were used with a modified and optimised protocol, described below. Kit reagents were prepared as per manufacturer's instructions. Where possible, DNA was extracted from a maximum of four buccal swabs obtained from a single dog. Using the QIAamp DNA Mini Kit, a maximum of two swabs were transferred into a 2 millilitre (mL) microcentrifuge tube with the addition of a lysis mix consisting of 400 microlitres (µL) phosphate buffered saline (PBS; Sigma-Aldrich Company Ltd., Dorset, UK), 400 µL AL lysis buffer (Qiagen) and 20 µL protease (Qiagen). Tubes were vortexed briefly followed by incubation at 56 °C for 15 minutes. DNA was precipitated by adding 400 µL ethanol (Thermo Fisher Scientific) to each tube and mixed by vortexing. The swabs and solution were transferred to a 5 mL syringe barrel held in a 15 mL conical tube (Falcon™) and centrifuged at 700 times gravity (×g) for 2 minutes. DNA was collected from the supernatant by applying 600 µL of sample solution to a QIAamp mini column and centrifuging at 6,000 ×g for 1 minute, discarding the flow-through. This was repeated until all the DNA-containing supernatant had passed through the column. DNA on the filter was purified using 500 µL AW1 wash buffer (Qiagen) and centrifuged at 20,000 ×g for 2 minutes. The flow-through was discarded and the addition of 500 µL AW1 wash buffer (Qiagen) and centrifugation was repeated. A final purification step was performed by adding 500 µL AW2 (Qiagen) wash buffer to each column and centrifuged for 20,000 ×g for 1 minute. The flow-through was discarded and the final purification step repeated. The filter column was placed in a fresh waste collection tube and centrifuged at 20,000 ×g for 3 minutes to dry the filter. Finally, the filter column was placed in a fresh 1.5 mL microcentrifuge tube and 100 µL AE elution buffer (Qiagen) was added to the filter. DNA samples were eluted by centrifuging at 6,000 ×g for 1 minute. A second 100 µL volume of AE buffer was

applied to each column and centrifuged at 6,000 ×g for 1 minute, resulting in a final volume of approximately 200 µL elute containing genomic DNA.

Where the QIAamp DNA Blood Midi Kit was used, four buccal swabs were suspended in a lysis buffer mix of 1,600 µL PBS (Sigma Aldrich), 1,600 µL AL lysis buffer (Qiagen) and 80 µL protease (Qiagen) in a 15 mL conical tube. Tubes were briefly vortexed and incubated at 56 °C for 30 minutes. DNA was precipitated by adding 1,600 µL absolute ethanol stored at -20 °C (Thermo Fisher Scientific) and the sample mixed by vortexing. The swabs and solution were transferred to a 5 mL syringe barrel held in a 15 mL conical tube and centrifuged at 700 ×g for 1 minute. Precipitated DNA was collected on a QIAamp midi column by applying 2.5 mL of the supernatant and centrifuging at 1,500 ×g for 3 minutes, discarding the flow-through. This step was repeated until all the supernatant had passed through the filter column. A final centrifugation at 2,000 ×g for 3 minutes was performed to ensure all solution had passed through the column, and flow-through discarded. DNA was purified by adding 2 mL AW1 wash buffer (Qiagen) and centrifuging at 4,300 ×g for 1 minute, discarding flow thorough. A second wash of 2 mL AW2 buffer (Qiagen) was added to the column and centrifuged at 4,300 ×g for 15 minutes, discarding the flow-through. DNA was eluted into a fresh 15 mL tube by adding 150 µL AE elution buffer (Qiagen) and centrifuging at 4,300 ×g for 1 minute. The elution step was repeated twice, resulting in a final volume of approximately 450 µL eluted gDNA.

DNA extraction from blood

Genomic DNA was extracted from up to 5 mL residual whole blood samples that were collected after clinical testing and preserved in EDTA (ethylenediaminetetraacetic acid) using an adapted version of the Nucleon BACC2 protocol (Tepnel Life Sciences, UK). Red blood cells were lysed by the addition of 25 mL of nucleon reagent A (see recipe in Chapter 2, section 2.12) to the residual blood and mixed by inversion. White blood cells were collected by centrifuging at 4,300 ×g for 10 minutes and the supernatant removed. Red blood cell lysis and centrifugation was repeated. White blood cell lysis was then performed by adding 2 mL nucleon reagent B (see recipe in Chapter 2, section 2.12) and the white blood cell pellet fragmented by shaking followed by incubation at 37 °C overnight. The cell lysis solution was transferred to a 15 mL conical tube, 800 µL of 5 molar (M) sodium perchlorate (see recipe in Chapter 2, section 2.12) was added and the tube mixed by inversion to achieve protein precipitation. A chloroform wash was performed by the addition of 2 mL chloroform (Thermo Fisher Scientific) followed by mixing by inversion for 4 minutes then centrifugation at 4,300 ×g for 5 minutes to separate aqueous phases. Avoiding contact with the middle protein layer, a sterile pipette was used to transfer the upper DNA-containing aqueous solution to a fresh 15 mL conical tube, leaving the bottom chloroform layer for disposal. The chloroform wash was repeated. As before, avoiding contact with the protein layer, a sterile Pasteur pipette was used to transfer the top DNA-containing aqueous solution to a

new 15 mL tube. DNA was precipitated by adding 5 mL cold absolute ethanol (Thermo Fisher Scientific) and slowly inverting the tube to encourage the DNA to spool. A sealed glass hook was used to hook out the DNA precipitate and left to air dry for 20 minutes. Finally, the precipitated DNA was washed off the glass hook into a 1.5 mL microcentrifuge tube using 300 µL 1× TE buffer (10 millimolar (mM) Tris, 0.1mM EDTA, Thermo Fisher Scientific) and left to dissolve at room temperature overnight. All extracted DNA samples were stored at -20 °C.

DNA quantitation and concentration

DNA concentration and purity were determined using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) and/or the Qubit Fluorometer with the Qubit dsDNA broad range (BR) Assay Kit (Invitrogen, Renfrewshire, UK). Where possible, DNA samples with low concentrations (less than 10 nanograms (ng)/µL) were concentrated using MultiScreen-PCR96 filter plates (Merck Millipore, Watford, UK) or Microcon −30 kDa centrifugal filter units with ultracel-30 membrane (Merck Millipore).

2.3 DNA amplification

Oligonucleotide primer and probe design

Gene annotations were obtained from the online genome browser Ensembl database across the human and/or dog genomes (Aken et al, 2016; Zerbino et al, 2018). All primers and custom probe-based qPCR assays for allelic discrimination assays were designed using Primer3 (Koressaar & Remm, 2007; Untergasser et al, 2012), listed in Appendix 4 for all assays, and obtained from Integrated DNA Technologies (IDT; Leuven, Belgium). Probes were PrimeTime ZEN double-quenched qPCR probes containing a 5' fluorophore, 3' Iowa Black® FQ (IBFQ) quencher and proprietary, internal ZEN™ quencher. For each custom probe-based qPCR assay, primers spanning amplicon sizes between 70–150 bp were designed and two probes: a 5′ HEX™ fluorophore was used to determine the reference/wild type allele and a FAM™ fluorophore (Carboxyfluorescein, 6-FAM) to label the alternate allele. All sequences of each designed primer pair were viewed through BLAST on Ensembl (http://www.ensembl.org/Multi/blastview) in the *Canis familiaris* genome CanFam3.1 (Sep.2011. Broad CanFam3.1/canFam3, Ensembl (www.ensembl.org; Dog release 89; Aken et al, 2016) to ensure alignment locations were correct for efficient amplification.

Oligonucleotide primer and probe dilution

All primers arrived as lyophilised primer pellets and were resuspended using ultrapure water (ddH_2O) to make 100 µM stocks. The 100 µM stock was used to make 20 µM working aliquots by combining 20 µL primer and 80 µL ddH_2O. Custom probe-based qPCR assays of premixed primers and probes (containing a primer pair and a probe labelled with FAM™ or HEX™ fluorophore) for allelic discrimination assays were resuspended in ddH_2O to a 10X or 40X assay (depending on the

amount of assay ordered). Equal quantities of the FAM™ and HEX™ labelled assays were combined to obtain a 10X or 40X probe mix.

Polymerase chain reactions (PCR)

The polymerase chain reaction (PCR) is a technique used to amplify segments of a DNA sequence and generate numerous copies of targeted DNA. The basic principle of PCR involves:

1. A denaturation step to separate double-stranded DNA to single strands.
2. An annealing step, usually between 55 °C and 60 °C, to allow binding of sequence-specific primers to the complementary sequence on the single-stranded DNA.
3. An extension stage at approximately 72 °C to allow the DNA polymerase to synthesise new strands of complementary DNA from the 3' end of the annealed primers.

These steps are typically repeated 30 to 40 times, allowing the DNA target to be exponentially amplified. HotStarTaq Plus DNA polymerase (Qiagen) was used for standard PCR and PrimeSTAR® GXL DNA Polymerase (Takara Bio Inc., Saint-Germain-en-Laye, France) was used to amplify long range templates by PCR. PCR reaction conditions are described in individual chapters.

Quantitative PCR (qPCR)

Quantitative PCR (qPCR) is a method used to determine the amount of a target sequence present in a sample. A custom probe-based qPCR assay was designed, as described above, flanking an exon junction to determine gene expression in tissue types of interest and compared to the housekeeping TATA box binding protein gene (*TBP*). Each reaction contained 4 µL Luna universal qPCR Master Mix (New England Biolabs (NEB) Ltd., Herts, UK), 1.2 µL ddH_2O, 0.8 µL of a 10X target assay mix and 2 µL complementary DNA (cDNA). qPCR was performed on a StepOnePlus™ Real-Time PCR system (Applied Biosystems, Loughborough, UK) using the following cycling conditions: 95 °C for 5 minutes followed by 40 cycles of 95 °C for 10 seconds and 60 °C for 30 seconds. A standard curve was generated using a 1:5 serial dilution of control cDNA and the coefficient of correlation (R^2) and amplification efficiency were determined. Results were analysed using Applied Biosystems StepOne Software v2.3. The R^2 value ascertains how well the data points lie on a single line, and how effective one value is at predicting another. Where R^2 is equal to 1, the value of Y (cycle threshold; Ct) can be used to accurately predict the exact value of x (e.g. the amount of DNA in the sample). If $R^2 > 0.99$, there is good confidence in correlating two values. Efficiency determines the sensitivity of a reaction and where efficiency is equal to 1 (or 100%), the fold increase will be 2 at each cycle. (Applied Biosystems, 2020). A qPCR efficiency from 90-110% is considered acceptable. qPCR efficiency is determined from a standard curve where a slope of -3.3 ± 10% reflects an efficiency of 100% ± 10% (Applied Biosystems, 2020).

A mean comparative Ct value was calculated between three biological replicates to determine expression levels of the targets. Target expression levels were normalised against *TBP* and compared by relative quantification. Normalised reporter (Rn) values were calculated as the ratio of the fluorescence of FAM™ dye divided by the fluorescence of ROX dye on the StepOnePlus™platform. Delta Rn (ΔRn) values were calculated using the Rn value of the experimental reaction minus the Rn value of the baseline signal generated by the StepOnePlus™ instrument.

Agarose gel electrophoresis

Agarose gel electrophoresis was used to confirm successful PCR amplification and determine DNA product sizes. Preparation of 1.5 % agarose gels was done using agarose (Bioline, London, UK), 1X TAE buffer (Tris-acetate-EDTA containing 40 mM Tris, 20 mM acetic acid and 1 mM EDTA, Thermo Fisher Scientific) and ethidium bromide (EtBr) solution (25 µg; Severn Biotech Ltd, Worcestershire, UK). Five µL loading buffer (see recipe in section 2.12) was added to 2 µL PCR product and all 7 µL loaded into the gel wells. Five µL of a 2-log DNA ladder (NEB; see recipe in section 2.12) was loaded alongside PCR products to measure DNA fragment sizes. DNA was separated by electrophoresis at 5 volts per centimetre (V/cm) for 40-60 minutes in a gel tank containing 1X TAE buffer (Thermo Fisher Scientific). The gels were visualised on an Alpha Innotech AlphaImager 3400 ultraviolet (UV) transilluminator (Alpha Innotech, California, USA) and gel images obtained with AlphaImager software. The separation ranges of bands were measured against the inserted 2-log DNA ladder (NEB) product image (Figure 2.1) measuring DNA fragment size ranges between 0.1-10 Kb.

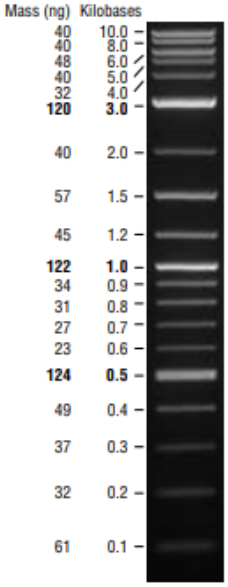


Figure 2.1: 2-Log DNA ladder visualised by EtBr staining on an agarose gel. Mass values indicated are for 1 µg/lane. Each separated band represents DNA fragment size within the range of 0.1 – 10 Kb. Image obtained from 2-log ladder product datasheet (NEB, UK).

2.4 Purification of PCR products

2.4.1 Filter plates

PCR products were purified to remove excess primers and reagents. The PCR product was made up to 100 µL using ddH_2O and transferred to a MultiScreen PCRµ96 filter plate (Merck Millipore) mounted on a vacuum manifold. A vacuum of approximately 15 inches Hg pressure was applied for 12 minutes to pass the solution through the filter wells. Resuspension of the purified PCR product was performed by the addition of 20 µL ddH_2O to each well followed by mixing on a rocking platform for 10 minutes. Purified PCR products were transferred to a 96-well PCR plate using a pipette for further use and stored at -20 °C.

2.4.2 Gel extraction

Where applicable, PCR products were purified following excision from an agarose gel using the Monarch® DNA Gel Extraction Kit (NEB), as per manufacturer's protocols. The desired band was excised from the gel using 6.5 x 1 millimetre (mm) gel cutting tips (Alpha Laboratories, Eastleigh, UK) with a P1000 pipette and transferred to a 1.5 mL microcentrifuge tube and the gel slice weighed. To the gel slice, 4 volumes of Gel Dissolving Buffer (NEB) (relative to the mass of the gel slice) were added and the sample incubated at 50 °C with regular vortexing until the gel slice was dissolved. For amplicons exceeding 8 Kb, an additional 1.5 volumes of ddH_2O was added after the gel slice was dissolved to mitigate the tighter binding of larger amplicons. The solute was transferred to a column inserted into a collection tube and centrifuged for 1 minute at 16,000 ×g, discarding the flow-through. The filter was washed twice by adding 200 µL DNA Wash Buffer

(NEB) and centrifuged for 1 minute at 16,000 ×g. The column was transferred to a fresh 1.5 mL microcentrifuge tube. Elution was achieved by the addition of 20 µL of DNA Elution Buffer (NEB) directly to the filter and left to stand at room temperature for 1 minute before centrifugation at 16,000 ×g for 1 minute.

2.5 Sanger sequencing

Single nucleotide variants (SNV) or small insertions and deletions (INDELs) were confirmed by Sanger sequencing.

Sanger sequencing reaction

PCR products to be used for Sanger sequencing were purified on a MultiScreen PCRµ96 filter plate (Merck Millipore), as described in section 2.4.1, and sequenced using the Sanger method. Each reaction contained 2 µL purified PCR product, 0.5 µL BrilliantDye Terminator v1.1 (Nimagen, Nijmegen, NL), 1.2 µL BrilliantDye Terminator Sequencing Buffer (5X) (Nimagen), 1.5 µL ddH_2O, and 1 µL 1.6 µM oligonucleotide primer (the forward or reverse oligonucleotide primer used in PCR amplification). Thermal cycling conditions were as follows: 96 °C for 30 seconds; 44 cycles at 92 °C for 4 seconds, 55 °C for 4 seconds, and 60 °C for 1 minute 50 seconds.

Isopropanol precipitation

Isopropanol precipitation of sequencing reaction products was performed to remove excess reagents. Firstly, 60 µL of 80% isopropanol (Thermo Fisher Scientific) was added to each well containing sequenced products, followed by centrifugation for 30 minutes at 2,600 ×g. The supernatant was discarded by immediate inversion and tapping of the plate upside down on tissue. A second addition of 100 µL of 60% isopropanol (Thermo Fisher Scientific) was added to each well and the plate centrifuged for 10 minutes at 2,600 ×g. The supernatant was discarded as before, and the plate left upside down on clean tissue. The plate was centrifuged upside down at 160 ×g for 1 minute, transferred the right way up and left covered at room temperature for 20 minutes to dry. Precipitated DNA was resuspended in 10 µL Hi-Di Formamide (Applied Biosystems). Sequencing products were separated on an Applied Biosystems (ABI) 3130xl Genetic Analyzer and raw sequence files generated for analysis.

Sequence data analysis

Sequence traces were assembled, analysed and compared using Pregap4 and Gap4 in the Staden software package (Staden et al, 2000).

2.6 Next-generation sequencing (NGS)

All NGS was performed using libraries generated from gDNA and sequenced on Illumina sequencing platforms. A standard NGS workflow is shown in Figure 2.2.

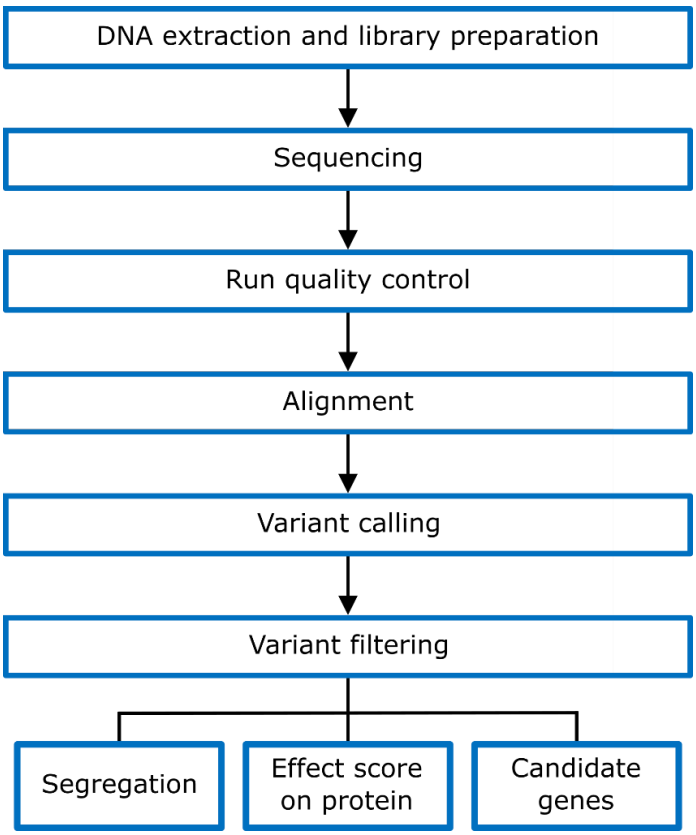


Figure 2.2: A standard workflow used in NGS analysis.

Library preparation

Libraries for amplicon sequencing were generated from 1 ng DNA using the Nextera XT library preparation reagents and protocol (Illumina, Cambridge, UK). Briefly, the Nextera transposome, a transposase-transposome complex, was used to tagment gDNA by simultaneously fragmenting and tagging DNA with adapter sequences (Figure 2.3A and Figure 2.3B). A 12-cycle PCR program amplified the tagmented insert DNA using read specific sequencing primers and added Index 1 (i7) and Index 2 (i5), common adapters (P5 and P7) (Illumina) and sequences required for sequencing cluster generation (Figure 2.3C). The addition of index adapters to both ends of the DNA enables dual-indexed sequencing of up to 96 DNA libraries in a single pool on an Illumina sequencing platform.

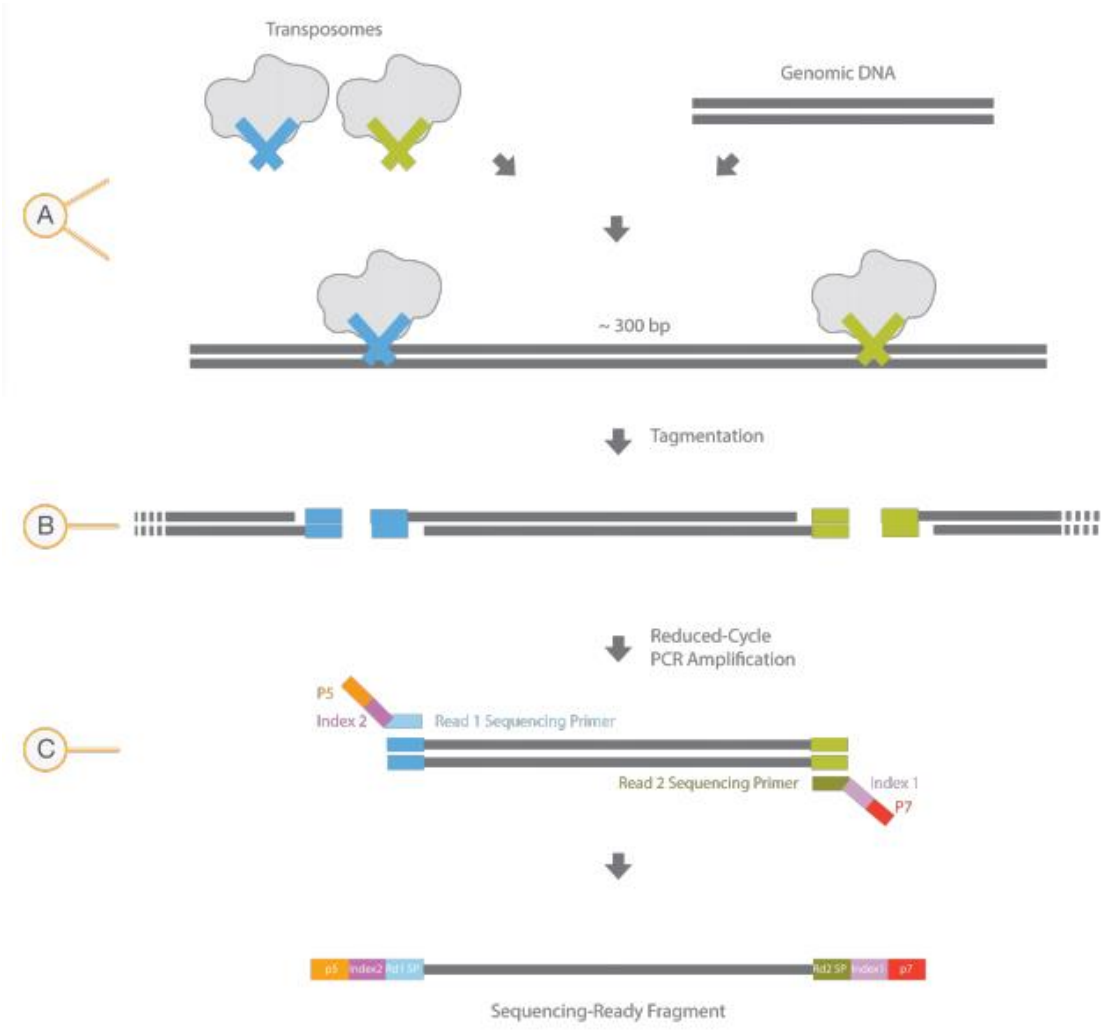


Figure 2.3: Schematic representation of Nextera XT DNA tagmentation process. **(A)** An engineered transposome with adapters is combined with template genomic DNA approximately 300-bp in size. **(B)** Tagmentation to fragment and add adapter sequences in one step. **(C)** A limited-cycle PCR anneals read sequencing primers (Rd1SP and Rd2SP) to add index adapter sequences (P5 and P7) to the end of each fragment and amplifies the insert DNA ready for dual-indexed sequencing (Illumina, 2017).

A clean-up step using AMPure XP beads (Beckman Coulter, High Wycombe, UK) used bead purification to purify amplified libraries. The quality of each individual DNA library was assessed by agarose gel electrophoresis and normalised using the Qubit Fluorometer with the Qubit dsDNA broad range or high sensitivity Assay Kit (Invitrogen), and then 3 µL of each individual library was pooled, thereby ensuring an equal representation of the library in the final pooled library. The pooled library was quantified using the KAPA library quantification kit, according to manufacturer's instructions (Kapa Biosystems, Massachusetts, USA) and diluted to a concentration of 2 nM. The pooled library was prepared for sequencing by denaturation using 5 µL DNA library and 5 µL 0.2 normality (N) sodium hydroxide (NaOH), incubating at room

temperature for 5 minutes followed by the addition of 990 µL pre-chilled HT1 buffer (Illumina) to the denatured library to achieve a concentration of 10 picomoles (pM). A total of 600 µL of 10 pM library was loaded into a MiSeq Reagent Kit v2 Nano cartridge (Illumina) to sequence on a MiSeq platform (Illumina).

Cluster generation

Following loading of the DNA library into a reagent cartridge, cluster generation was performed by the sequencing platform to generate clonal clusters on the flow cell in preparation for sequencing. Here, each single-stranded DNA fragment binds randomly to oligonucleotides on the surface of the flow cell, facilitated by the adapters at each end of the DNA fragment added in the library preparation stage (Figure 2.4). In a process called bridge amplification, single-stranded molecules flip over to hybridise to adjacent primers and unlabelled nucleotides and polymerase enzyme initiate solid-phase amplification, generating double stranded fragments on the flow cell surface. This bridge amplification process is repeated until each bound fragment is amplified into a clonal cluster creating several million dense clusters in each channel of the flow cell. The double-stranded bridges are denatured leaving single-stranded templates anchored to the surface of the flow cell. Reverse strands are cleaved and washed away, leaving single-stranded DNA molecules attached to the flow cell surface. Free 3' ends of single-stranded DNA are blocked to prevent unwanted DNA priming.

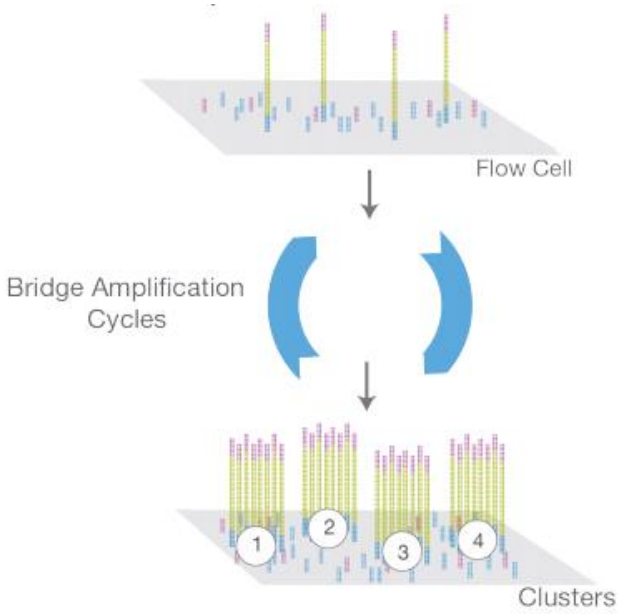


Figure 2.4: Illustration of cluster amplification in next-generation sequencing. DNA fragments from the loaded DNA library are hybridised to the flow cell surface and amplified into a clonal cluster through bridge amplification (Illumina, 2017).

Sequencing

Sequencing primers hybridise to the single-stranded DNA molecules and other sequencing reagents, including four fluorescently labelled nucleotides (A, C, G, T) and DNA polymerase are added to the clusters to initiate sequencing. The polymerase adds the first nucleotide to the primer, using the single-stranded DNA as a template. The identity of this base is determined by imaging of the flow cell flowing laser excitation, capturing the emission wavelength of fluorescence from each cluster. This cycle of incorporating the labelled nucleotides, primers and DNA polymerase followed by laser excitation and image capturing is repeated '*n*' times to create a sequence read with a length of '*n*' bases. Reads generated from each cluster are exported to an output text file in FASTQ format for data analysis. A schematic of each step in the sequencing process is shown in Figure 2.5

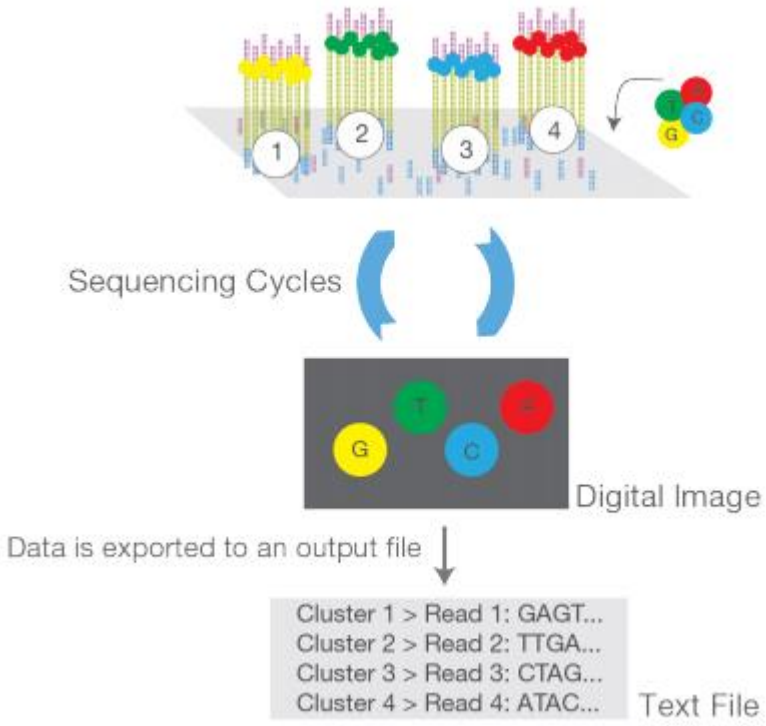


Figure 2.5: Sequencing of single-stranded DNA molecules following cluster amplification is initiated by the addition of sequencing reagents, including fluorescently labelled nucleotides A, C, T and G. Sequencing cycles commence and a digital image of the flow cell is captured, with the emission of each cluster recorded to identify the base. Reads generated from each cluster are exported to an output text file for data analysis (Illumina, 2017).

Data analysis

Raw sequencing data files were received in FASTQ file format and were processed through the AHT in-house pipeline, based on the Genome Analysis Toolkit (GATK) Best Practices workflow (DePristo et al, 2011; Van der Auwera et al, 2013). This NGS pipeline uses a number of analytical tools to align sequence files, assess the quality of the sequencing reads to retain high quality data, identify sequence variants, make variant effect prediction calls and enable filtering steps to be performed; which can be tailored to each study depending on the disease mode of inheritance. Alignment of short sequence reads to the canine reference sequence (CanFam3.1) is carried out using the BWA (Burrows-Wheeler-Alignment) tool that generates a Sequence Alignment/Map (SAM) file (Li & Durbin, 2009). Picard tools (http://broadinstitute.github.io/picard) sorts the aligned reads in the SAM file based on chromosomal coordinates and converts it to a Binary Alignment/Map (BAM) format whilst adding sequencing read group information to the BAM file. BAM files can be manually visualised using Integrative Genomics Viewer (IGV) software (Robinson et al, 2011; Thorvaldsdottir et al, 2013) at this point if a region of interest or candidate genes are to be interrogated.

The Picard 'ValidateSamFile' tool reports on the validity of the BAM file relative to the SAM file and writes any errors caused by faulty alignment or improper formatting to an output file. The SAMtools bioinformatics package is a set of utilities that allows alignment manipulation of SAM and BAM files (Li et al, 2009). SAMtools was used to remove PCR duplicate reads, retaining high quality read pairs, in addition to indexing the reference FASTA files and read BAM files. GATK is an open-source Java framework platform providing tools for NGS analysis including variant calling and quality control of high-throughput NGS data (McKenna et al, 2010). GATK is incorporated into the in-house pipeline to facilitate variant calling. Variant calls were made using HaplotypeCaller (GATK); base quality score recalibration, INDEL realignment and duplicate removal were performed (McKenna et al, 2010). SNP and small INDEL discovery were performed using standard hard filtering parameters or variant quality score recalibration according to GATK Best Practices recommendations (DePristo et al, 2011; Van der Auwera et al, 2013). Genomic Variant Call Format (gVCF) files from multiple canine WGS were merged by CombineGVCFs (GATK) into a multi-sample Variant Call Format (VCF) file. Cross-genome analysis was conducted using Variant Effect Predictor (VEP) provided by Ensembl on a merged VCF file.

Whole genome sequencing (WGS)

Whole genome sequencing (WGS) utilises NGS technology to provide a high-resolution, base-by-base view of an entire genome for further analysis. A large volume of data is obtained where all exonic and non-exonic regions of the genome are captured, subject to sufficient coverage over GC-rich regions or repetitive structures in the DNA due to incorrect assembly over these difficult regions.

To select buccal cheek swab samples for WGS, the extent of bacterial DNA contamination in each sample was determined. Libraries were generated from 1 ng DNA using the Nextera XT library preparation reagents and protocol (Illumina; described above), and sequenced on the MiSeq platform using the MiSeq Reagent Kit v2 Nano cartridge (Illumina). Paired-end reads 50-bp in length were generated. Reads were aligned to the canine reference genome using BWA (Li & Durbin, 2010) and the percentage of reads aligned were determined using the "flagstat" command in SAMtools (Li et al, 2009). Samples with > 80% alignment (i.e. < 20% bacterial content) were considered suitable for WGS. Using this method, the number of reads generated per sample (14,669-102,068) provides less than 1X coverage, but is sufficient to give an indication of the level of non-dog DNA contamination of the DNA sample.

Genomic DNA samples selected for WGS were normalised to 25 ng/µL, and 1,000 ng was sent for sequencing, outsourced to Edinburgh Genomics, UK. Illumina sequencing of a TruSeq Nano library on a HiSeq X sequencing platform (Illumina) generated a dataset of approximately 30X coverage of the dog genome. Sequencing data were analysed as described above.

***De novo* assembly**

To perform *de novo* assembly over regions of interest where coverage was low in WGS data or the region was unannotated in the canine reference sequence, the presence of DNA sequence from an Illumina read visible in IGV in the region of interest at either end of the gap was queried in the FASTQ files using an in-house script. The Linux command-line utility 'grep' implemented in the script searched for a chosen run of nucleotides (20-30 bp), in both the forward and reverse orientation, that were present in all of the sequencing reads in the FASTQ file. The perl script used for this de novo assembly is shown in Appendix 5. An output file was generated that contained a list of the FASTQ Illumina reads where the searched run of nucleotides was present. This was repeated, each time querying a new run of nucleotides further upstream or downstream of the previous sequence queried to walk across the sequence 'gap'. Each sequence read extracted from the FASTQ file was converted to FASTA format and reads were assembled into contigs using Pregap4 and Gap4 in the Staden software package (Staden et al, 2000).

2.7 Exclusion of known retinal mutations

Previously published retinal mutations were screened in two to three cases prior to further sample recruitment and investigation in each breed to exclude them as causal variants of PRA (Appendix 2). SNVs or small INDELs were genotyped using a genotyping-by-sequencing method, described below. Large INDELs or variants within repetitive regions were genotyped using PCR amplification, followed by amplified fragment length polymorphism (AFLP) analysis or visualisation on an agarose gel using gel electrophoresis.

To perform genotyping-by-sequencing, primer pairs were designed spanning each mutation to amplify DNA fragments between 110–130 bp in size. Primer pairs (Appendix 4, Table A1A) were pooled and a multiplex PCR was performed. A primer mix was made by combining 4 µL of each 100 µM stock primer with 468 µL of ddH_2O, achieving a final concentration of 0.2 µM of each primer. Each PCR reaction comprised 5 µL primer mix, 2.27 µL deoxyribonucleotide triphosphate (dNTP) mix (Thermo Fisher Scientific) at a final concentration of 0.2 mM, 1.7 µL of 10X PCR buffer (Qiagen), 1.02 µL $MgCl_2$ (Qiagen) at a final concentration of 1.5 mM, 0.3 µL HotStar Taq (5 units/ µL) (Qiagen) and 4.71 µL ddH_2O. Thermal cycling conditions for the multiplex PCR are listed in Appendix 4. Purification was carried out after each thermal cycling reaction using AMPure XP beads (Beckman Coulter), according to manufacturer's instructions, and using 1:1.75 ratio of beads to DNA-containing solution.

Sequencing libraries were prepared as explained in section 2.6. Five µL of each sample DNA library was pooled and quantified using a KAPA library quantification kit, according to the manufacturer's instructions (Kapa Biosystems). The final library was diluted to 15 pM and loaded into a 150-bp v3 kit cartridge (Illumina) for single-read sequencing on a MiSeq sequencing platform (Illumina). FASTQ files were aligned to the canine genome assembly CanFam3.1 using

BWA, producing BAM files. BAM files were visualised in IGV (Robinson et al, 2011; Thorvaldsdottir et al, 2013).

PRA-associated variants that were not included in laboratory screening methods prior to undertaking additional genetic analysis (e.g. those within regions difficult to amplify or identified after the samples were selected for WGS) were looked for in WGS data of each case to exclude as causal for each breed (Appendix 2).

2.8 Variant filtering

Effect scores

An in-house analysis pipeline generated an effect score for each variant based on the predicted effect of the variant on the protein sequence i.e. scored whether it is predicted to be deleterious. This allows the ranking of all variants. All effect scores and their assignment to sequence ontology terms can be found in Appendix 6, as termed by the Ensembl database, and the locations of each display term relative to the transcript structure is illustrated in Figure 2.6. As an initial step, variant filtering retained only variants with a high effect score (4 and 5), including those resulting in premature start/stop codons, splice site variants, nonsense and missense variants, frameshift variants, and in-frame INDELs.

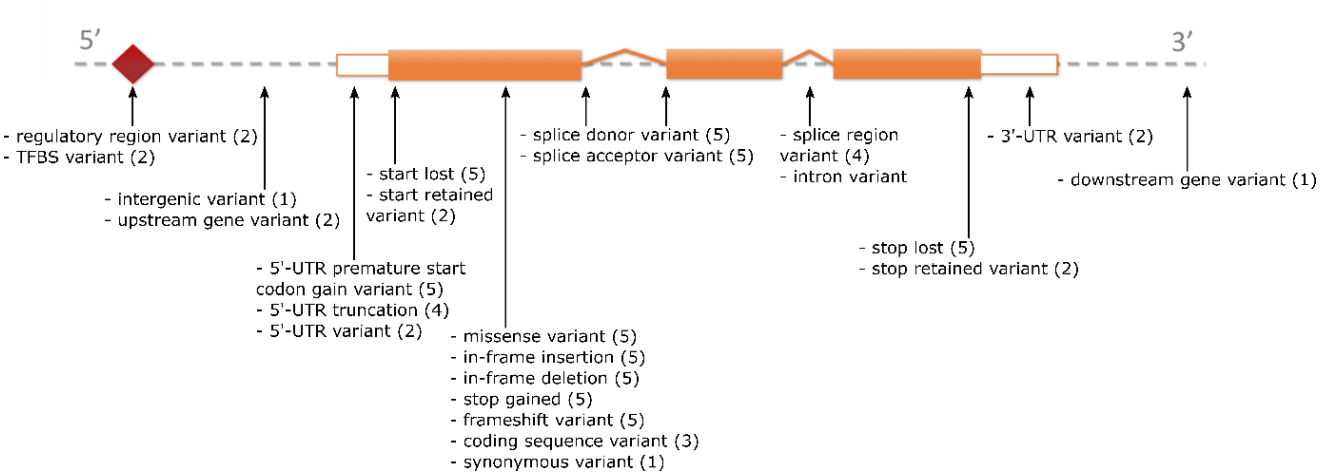


Figure 2.6: Effect score sequence ontology terms. A diagram showing the locations of sequence ontology terms used to determine effect scores to the relative transcript structure (EMBL-EBI, 2020). Assigned effect score for the sequence ontology term is in brackets. A full list of terms can be found in Appendix 6.

Segregation scores

To further aid variant filtering, variants in cases were compared against control dogs in the analysis and were assigned in-house segregation scores, calculated as follows:

Segregation score = Number of samples – Discordant samples

Where the 'Number of samples' is the count of all the individuals used in the analysis (i.e. the number of cases plus the number of controls), and 'Discordant samples' is the number of those individuals in which the genotype at a particular locus does not match the genotype expected based on the mode of inheritance.

A high segregation score indicates that the variant segregates with disease status. The highest possible segregation score is the total number of individuals used in the analysis i.e. in an analysis of 1 case versus 199 controls, the highest segregation score assigned is 200. Where a control individual shares the same allele at a given position as the affected individual under investigation, a lower segregation score is assigned due to the reduced likelihood of a causative variant being present in a control dog. Depending on the disease model under investigation, segregation scores can be used with effect scores to prioritise candidate variants and filter out variants that are likely to be polymorphic, benign, or non-pathogenic.

OMIM database

The Online Mendelian Inheritance in Man (OMIM) database (OMIM, 2020) is an online catalogue of genes implicated in hereditary disorders in humans. The following list of keywords related to PRA were compiled and cross-referenced through the OMIM database to extract genes within a candidate list from NGS analysis that were associated with any keywords in the PRA keyword list: retina, progressive retinal atrophy, retinitis pigmentosa, rod-cone dysplasia/rod cone dysplasia, cone-rod dysplasia/cone rod dysplasia, cone degeneration, retinal degeneration, rod dysplasia, photoreceptor, retinal dystrophy, retinal pigment epithelium, nyctalopia, bone spicules, fundus, blindness, Leber congenital amaurosis, Bardet-Biedl syndrome/Bardet Biedl syndrome, chorioretinal atrophy, chorioretinal degeneration, cone dystrophy, cone-rod dystrophy/cone rod dystrophy, rod-cone dystrophy/rod cone dystrophy, congenital stationary night blindness, macular degeneration, ocular-retinal developmental disease/ocular retinal developmental disease, optic atrophy, Usher syndrome, retinopathy, cilium, ciliopathy. This highlighted genes with previous evidence of association with ocular phenotypes and allowed prioritisation of high impact variants based on gene function.

Dog Biomedical Variant Database Consortium

The AHT Canine Genetics group is a member of the Dog Biomedical Variant Database consortium (DBVDC) (Jagannathan et al, 2019), where members are invited to submit canine genomes in the form of FASTQ files that are processed into a multi-sample VCF file to be shared with other members of the consortium. The multi-sample VCF file was used for additional filtering. Although there is usually no phenotype or clinical information submitted with genetic information, members can contact owners of genomes for further information if a dog harbours a variant of interest to aid in inclusion/exclusion of candidate variants in genetic investigations.

2.9 Variant screening

Screening methods were chosen based on those appropriate for the variant in question. Typically, small INDELs were assayed using AFLP analysis and SNVs using allelic discrimination assays.

Amplified fragment length polymorphism (AFLP) analysis

AFLP analysis was used for genotyping small INDELs that caused a change in the size of an amplicon. PCR products were generated used a specific primer pair, of which one incorporated a FAM™ fluorescence (Carboxyfluorescein, 6-FAM). Capillary electrophoresis on a 3130xl genetic analyser (Applied Biosystems) was used to separate PCR products and determine allele sizes. This technique enables amplicons differing by a single base-pair to be separated and identified, distinguishing between wild type and alternate alleles thus also enabling the identification of heterozygotes. Each well contained 1 µL PCR product, 9.59 µL Hi-Di formamide (Thermo Fisher Scientific,) and 0.41 µL GeneScan™ 400HD ROX or 500 LIZ dye size standards (Thermo Fisher Scientific). GeneMapper software (Applied Biosystems) was used to interpret data.

Allelic discrimination assays

Allelic discrimination assays were typically employed for genotyping of SNVs at a given position in the DNA, using allele-specific fluorescently labelled probes. Forward and reverse primers flanking the variant and two allele-specific probes complementary to the two alleles at the variant position were designed and prepared as described in section 2.3. Each allele-specific probe is labelled with a fluorescent reporter dye (fluorophore FAM™ or HEX™) on the 5' end and a quencher dye (IBFQ) on the 3' end. When the reporter dye is in close proximity to the quencher, the quencher absorbs the energy emitted by the reporter dye causing a reduction in fluorescence. Where the allele complementary to the probe is present, probe hybridisation occurs causing the quencher to be cleaved by the 5' to 3' exonuclease activity of DNA polymerase. Separation of the reporter dye from the quencher emits fluorescence. A mismatch between probe and template DNA reduces the efficiency of the probe hybridisation and cleavage, and therefore reduces the reporter signal of that probe (Livak, 1999). Alleles can be distinguished by determining which reporter dye is fluorescing, with an increase in both reporter signals indicating heterozygosity. A real-time PCR is used to amplify 2µL template DNA in 0.2 µL of a 10X or 40X assay mix, 4 µL Luna universal qPCR Master Mix (NEB) and 1.8 µL ddH_2O on a StepOnePlus™ Real-Time PCR system (Applied Biosystems). Results were analysed using Applied Biosystems StepOne Software v2.3.

2.10 Genome-wide association study (GWAS)

A GWAS approach was used to investigate PRA in the LA using a case-control study design. A flow chart illustrating the steps performed in this GWAS approach is shown in Figure 2.7.

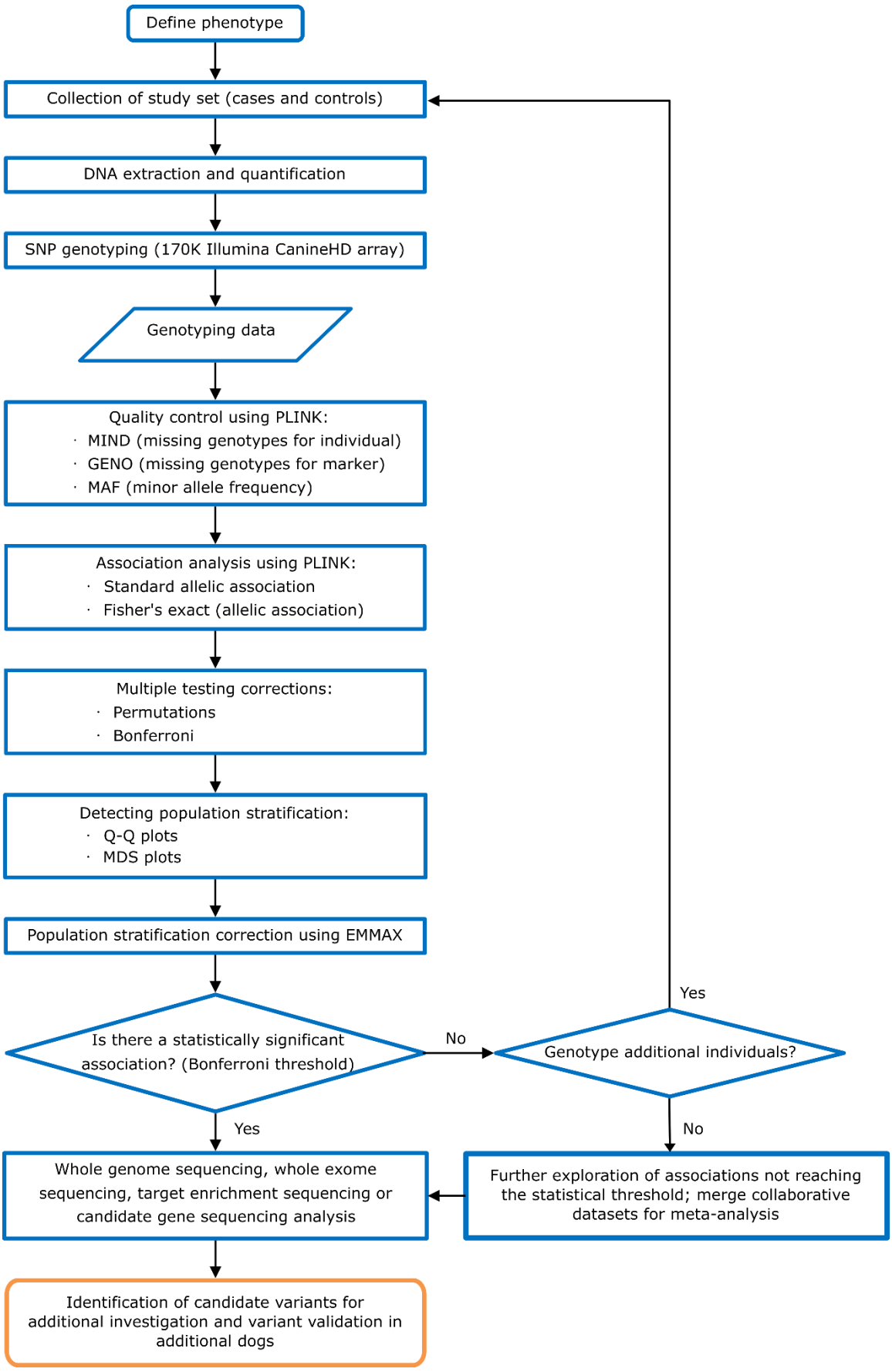


Figure 2.7: A flowchart illustrating a GWAS study design.

SNP genotyping

Five to twenty micrograms (μg) of DNA was sent externally to GeneSeek (Neogen Europe, Ayr, UK) for SNP genotyping, carried out using the 170K Illumina CanineHD array (Illumina, San Diego, USA) comprising 172,115 SNPs. Data were returned in GenomeStudio (Illumina) format.

Quality control and genotyping analysis

GWAS analysis was conducted using the free open-source toolset PLINK v1.90 (http://zzz.bwh.harvard.edu/plink/) (Purcell et al, 2007), adapted for use in canine GWAS at the AHT. An in-house Perl script termed 'PLINKDOG' added the following options in effect to each PLINK command: --noweb --dog and --allow-no-sex.

Genotyping data were exported from GenomeStudio to PED and MAP files. The PED file is a tab delimited text file with six mandatory columns: 'Family ID', 'Individual ID', 'Paternal ID', 'Maternal ID', 'Sex' and 'Phenotype'. Each line in a MAP file describes a single marker and four columns exist: 'chromosome', 'rs# or SNP identifier', 'genetic distance' (morgans) and 'base-pair position' (bp units). Instead of assigned Family identity (ID) numbers, each individual was numbered in the order they appeared in the file as this was not needed for the GWAS. Individual identity (IID) numbers were assigned to each sample from the raw data files using the AHT database number. Paternal and Maternal ID were not necessary in this study and therefore were assigned 0 as missing. Sex was indicated by 1 (male), 2 (female) or 0 (missing); set as 0 as default. By default, the phenotype was missing in the raw PED file and needed to be combined with PED and MAP files from GenomeStudio (Illumina) before analysis could be conducted. A phenotype file was generated including the family ID (FID) number (assigned by the data file), an IID number and a phenotype number: 2 (cases), 1 (controls). This was achieved by using the following PLINK command, where X is the file name prefix:

--plinkdog –-file X –-pheno X.pheno.txt –-keep X.pheno.txt –-recode –-out X

Genome-wide association mapping was performed by standard allelic association analysis using the PED and MAP files using the following command:

--plinkdog –-file X –-out X –-assoc –-maf 0.05 –-geno 0.03 –-mind 0.1

This command also includes an initial quality control (QC) filtering for minor allele frequency (MAF), missing genotypes per SNP (GENO) and missing genotypes per individual (MIND). SNP markers with an allele frequency of less than 5% (MAF 0.05) were excluded, including monomorphic SNPs, as they were considered too rare to be informative i.e. where differences between the frequency of the allele in cases and controls could be due to chance. SNP markers with a genotyping call rate of less than 97% (GENO 0.03) were excluded, based on previous analysis within the AHT Canine Genetics research group that showed this threshold was

appropriate to preserve high quality data for these arrays. By default, individuals within the dataset with more than 10% missing genotypes were omitted (MIND 0.1).

The above script performs a standard case-control association analysis comprising an unadjusted allelic chi-squared (χ^2) test for association (1 degree of freedom). This computes the frequency of both alleles in cases and controls and performs a χ^2 test, obtaining raw p-values (p_{raw}). Data from a GWAS are typically reported using a Manhattan plot representation, plotting each individual SNP using the negative log ($-\log_{10}$) of individual p-values against SNP genomic location. SNPs are positioned across the x-axis according to chromosomal position, ordered by ascending genomic position (chromosome 1-X). The y-axis represents the $-\log_{10}$ of the SNP's p-value for association with the phenotype. This depicts the strength of associations across the genome, positioning SNPs with the lowest p-values, and therefore highest association with disease, higher up on the graph. Although the most associated SNP from a GWAS analysis is unlikely to be a causal variant, a Manhattan plot highlights genomic regions of interest that can be further interrogated.

The type I error, also known as the significance level or false positive result, is the probability of rejecting the null hypothesis when it is true (Clarke et al, 2011). This incorrectly ascribes scientific significance to a statistical test, i.e. assigns significance to a SNP when it is not significant and therefore not associated with a disease/trait of interest, which is inappropriate in many cases. . Type 2 errors, or false negative results, can occur if the correction for multiple comparisons is too stringent or if the power is inadequate (Johnson et al, 2010). To reduce the risk of type 1 and type 2 errors in canine GWAS analyses, dogs of the same breed should be used considering demographic history. Selecting individuals from a varied population, i.e. limiting the number of closely related individual dogs within the study cohort, will also help minimise the incidence of false positives. In contrast to human GWAS, fewer SNPs are required for canine GWAS to identify an association and therefore the risk of type 1 errors is lower. In addition, the canine genome shows higher levels of LD. With these long stretches of LD, multiple SNPs will have similar statistical significance, which may be correlated with one another, and introduce false negative results by reducing the power of true disease associated markers, increasing the risk of type 2 errors. As hundreds of thousands of statistical tests are performed in GWAS analysis due to the extensive number of SNP markers genotyped (known as multiple testing), many of these tests will result in a significant p-value ($< p = 0.05$) by chance. Multiple testing correction is required to adjust for the simultaneous testing of multiple hypotheses. Genome-wide significance was defined using the Bonferroni correction method to identify significant associations. This is considered the most conservative method for selecting a p-value threshold as it assumes independence among the association tests (SNPs). A Bonferroni corrected significance threshold was defined by dividing the p-value 0.05 by the number of SNPs tested in the GWAS that remained after QC filtering, i.e. the number of statistical tests, and represented as a $-\log^{10}$ value on the Manhattan

plot. The permutation test, considered the gold standard in adjusting groups of correlated tests for multiple testing, accurately takes into account the correlation structure of the genome (Joo et al, 2016) and provides evidence that a positive hit from a significant p-value is not likely to be a false positive. To adjust for multiple testing, the PLINK Max(T) permutation procedure was employed, performing 100,000 permutations (--mperm) (Purcell et al, 2007).

Population stratification refers to the presence of subpopulations in a group of individuals, most likely due to divergent ancestry or relatedness amongst samples. If cases and controls are poorly matched based on their ancestry or relatedness or unbalanced between cases and controls, allele frequencies could differ between cases and controls because they are not representative of each other rather than due to an association with disease (Rogers & Weiss, 2017), potentially giving rise to false positive results. The only difference that should exist between cases and controls in the study is their phenotype for the disease under investigation, although canine researchers tend to favour 'super controls' that are over the typical age of onset for the disease under study. Diagnostic plots based on GWAS data can be used to assess the presence of population substructure. Two main plots are used to visualise GWAS data: multi-dimensional scaling (MDS) plots and quantile-quantile (Q-Q) plots. To perform MDS analysis, the following PLINK command was executed:

--plinkdog –-noweb --dog –-allow-no-sex –-nonfounders -–file X –-make-founders –--cluster –-mds-plot 2 -–out X

MDS plots can be used to identify individual outliers in a population due to a difference in ancestry or where cases and controls are not representative of each other. Each individual in the PLINK MDS output file was assigned a phenotype number of either 1 (control) or 2 (case). Data from this file were then plotted as scatter graphs with the first dimension (covariate 1) on the x-axis and the second dimension (covariate 2) on the y-axis with each point representing an individual, either a case or control.

Q-Q plots detect inflation of GWAS test statistics by enabling a graphical representation of how much the observed p-values deviate from the expected p-values from a hypothetical χ^2 distribution when plotted against one another. The overall degree of stratification, known as the inflation factor (lambda, λ), is calculated based on comparing the expected and observed median p-values; in the absence of stratification lambda is very close to 1. This occurs when the observed values correspond to the expected values until significant true observed associations are identified at the top end scale where the two should diverge. Where the majority of observed values begin to deviate from the expected throughout, this indicates that many moderately significant p-values are reported as more significant than expected under the null hypothesis and is likely to be due to the presence of population stratification (Ehret, 2010). Conversely, when only

a few of the most significant observed values deviate from the expected values, it is an indication that population stratification is not a factor and the association is real. Q-Q plots were generated as follows:

1. The $-\log_{10}$ p-value was generated for each observed p-values from the PLINK association output file.
2. These $-\log_{10}$ observed p-values were sorted from largest to smallest and assigned a rank from 1 to the total number of SNPs i.e. ranked 1 to 108,263.
3. An expected p-value was calculated by subtracting 0.5 from the rank number and dividing by the total number of SNPs tested e.g. for rank 1 value: (1 minus 0.5) divided by 108,263.
4. A $-\log_{10}$ p-value was generated for each expected p-value.
5. The expected (x-axis) and observed (y-axis) $-\log_{10}$ p-values were plotted on a graph. The baseline representing lambda as 1 was generated by plotting the expected p-value for each marker against itself on the x-axis and y-axis.
6. The genomic inflation factor (lambda) was generated by calculating the median for the observed and expected values, and dividing the observed median by the expected median.

Population stratification and sample relatedness in the GWAS was corrected for using Efficient Mixed-Model Association eXpedited (EMMAX) (Kang et al, 2010). This is a statistical test performed separatley to PLINK analysis in order to exclude false-positive associations of SNPs with the disease phenotype by correcting the PLINK results. The association analysis performed with PLINK corrects for population structure but does not account for relatedness in the sample. Therefore, EMMAX is used to correct the PLINK result to account for both ancestry and population structure.

2.11 *In silico* tools

The Ensembl genome browser (www.ensembl.org; Dog releases 87-98) (Aken et al, 2016) and UCSC genome browser (http://genome.ucsc.edu/) (Kent et al, 2002) were used to obtain canine reference genome sequence (Sep.2011. Broad CanFam3.1/canFam3) to interrogate regions. All features, such as genes, transcripts, exons and proteins, that have been annotated and mapped in the genome are given unique identifiers. Where genes have been annotated with known function, information or orthology, a gene name (gene symbol) is also assigned. Where a gene name has not been assigned in the canine genome, but sequence annotation and chromosomal position is known, an Ensembl stable identifier is assigned. In the dog, this is given the prefix 'ENSCAF' followed by a unique identifier number. The Ensembl BioMart data mining tool (Ensembl Dog release 87) (Aken et al, 2016) was used to search for human orthologues of unassigned canine gene names (i.e. querying canine ENSCAF numbers) to obtain human gene names.

Mutation Analyzer (NCBI Mutation Analyzer, 2017), PolyPhen-2 (Adzhubei et al, 2010) and SIFT (Kumar et al, 2009) were used to predict potential impact of amino acid substitutions on the structure and function of proteins. Mutation Analyzer uses a BLOSUM62 scoring matrix to compare amino acid substitutions, reporting residues that substitute frequently as positive scores (coloured green) and those that substitute rarely as negative scores (coloured red). PolyPhen-2 calculates the probability that a given mutation is damaging, reporting estimates of false positive rates (the chance that the mutation is classed as damaging although it is non-damaging) and true positive rates (the chance that the mutation is classed as damaging when it is indeed damaging) (Adzhubei et al, 2010). PolyPhen-2 also reports qualitative classification of a mutation as benign, possibly damaging or probably damaging. A SIFT score is a normalised probability of observing the new amino acid at that position, and ranges from 0 to 1 where a value between 0 and 0.05 is predicted to affect protein function (Sim et al, 2012). Putative promoter regions were predicted using Gene2Promoter (Genomatix software suite, 2017) and PromoterInspector (Scherf et al, 2000) and transcription factor binding sites (TFBSs) using MatInspector (Cartharius et al, 2005).

2.12 Buffers and solutions

Nucleon reagent A

To make 1 litre (L), 1.576 grams (g) Trizma Hydrochloride (Sigma-Aldrich), 109.53 g Sucrose (Sigma-Aldrich), 1.017 g $MgCl_26H_2O$ (Sigma-Aldrich), 19 mL Triton X-100 (Sigma-Aldrich) and 900 mL ddH_2O were combined. The pH was adjusted to 8.0 with 40% NaOH (Sigma-Aldrich) and the solution autoclaved before use.

Nucleon reagent B

To make 1 L, 63 g Trizma Hydrochloride (Sigma-Aldrich), 22.3 g EDTA (Sigma-Aldrich), 8.8 g sodium chloride (NaCl; Sigma-Aldrich) and 800 mL ddH_2O were combined. The pH was adjusted to 8.0 with 2M NaOH (Sigma-Aldrich) and the solution volume adjusted to 1 L then 10 g Sodium dodecyl sulphate (SDS; Sigma-Aldrich) added.

Sodium perchlorate solution (5M)

To make 1 L, 612.2 g sodium perchlorate (5M) was dissolved in ddH_2O to a volume of 1 L and filtered using a Nalgene sterile single use vacuum driven filter unit (Thermo Fisher Scientific).

Loading buffer

Ten mL of glycerol was combined with 4 mL 500 mM EDTA, 1 mL 1 M TRIZMA® Base and 1 mL bromophenol blue (all from Sigma-Aldrich). The solution was diluted 1:5 with 1X TAE buffer.

DNA size marker "2-Log DNA ladder"

Forty µL of 2-Log ladder (NEB) was combined with 120 µL loading buffer and 240 µL 1X TAE buffer.

Cell culture growth medium

The following were added to a 500 mL bottle of Dulbecco's Modified Eagle Medium (DMEM) with 4.5 g glucose: 10% fetal bovine serum, 2 mM penicillin-streptomycin mix and 2 mM L-glutamine (all from Invitrogen) and stored at 4 °C for use within six weeks.

Lysogeny broth (LB)

To make 500 mL, 5 g Tryptone, 2.5 g yeast extract and 2.5 g NaCl (all from Sigma-Aldrich) were added to 450 mL ddH_2O and stirred on an electronic stirring plate using a stirring flea. Once dissolved, the volume was made up to a total of 500 mL and autoclaved. If antibiotic was to be added, this was done following autoclaving once the liquid had reached a temperature approximately under 50 °C to prevent antibiotic degradation and stored at 4 °C for use within six weeks.

Lysogeny broth (LB) agar

To make 500 mL, 5 g Tryptone, 2.5 g yeast extract and 2.5 g NaCl (all from Sigma-Aldrich) were added to 450 mL ddH_2O and mixed by inversion by hand. Following this, 7.5 g Bacteriological agar (Sigma-Aldrich) was added and the solution stirred on an electronic stirring plate using a stirring flea. Once dissolved, the volume was made up to a total of 500 mL and autoclaved. If antibiotic was to be added, this was done following autoclaving once the liquid had reached a temperature approximately under 50 °C to prevent antibiotic degradation. Following the addition of antibiotic, plates were stored at 4 °C and used within six weeks.

Chapter 3 PRA in the Giant Schnauzer

3.1 Introduction

The Giant Schnauzer (GS) dog is the largest of three recognised sizes within the Schnauzer group: Giant, Standard, and Miniature, and is believed to have originated in Germany. Although the precise origins are unclear, it is thought the breed was developed using a variety of breeds including the Standard Schnauzer, Rottweiler, Great Dane, German Shepherd, and perhaps Bouvier De Flandres (American Kennel Club, 2007). Today, the three breed sizes are recognised as distinct breeds by The Kennel Club, UK and do not interbreed. A total of 282 GS were registered with The Kennel Club, UK in 2019 and over the last decade, this registration total has not changed significantly, with an average of 238 dogs registered per year since 2010 (The Kennel Club UK, 2020).

Currently, the GS is recognised by the BVA/KC/ISDS eye scheme to suffer from hereditary cataracts (HC). In 2018, 912 GS dogs underwent a BVA/KC/ISDS eye examination of which 16 were diagnosed as affected HC (British Veterinary Association, 2018). To date, PRA has not been listed on the current or previous versions of the eye scheme for the GS and is not considered to be seen frequently enough to warrant routine screening of PRA in this breed.

This chapter describes the identification of a previously unidentified PRA-associated variant in perhaps the first reported PRA cases in the UK GS population. The gene implicated has not previously been associated with retinal degeneration in any species. Although the small number of cases suggests this is a newly emerging form of PRA in the breed, identification of the genetic cause will enable further screening in the breed to confirm or exclude this hypothesis and improve understanding of the aetiology of this form of PRA.

A manuscript describing this study was accepted for publication following peer-review in the Canine Genetics Special Issue of the journal, *Genes* in May 2019. Rebekkah Hitti-Malin performed all laboratory and in silico experiments, formal analysis, data curation, validation and manuscript preparation. The manuscript was reviewed by all authors, with editing contributions from Dr Cathryn Mellersh, Dr David Sargan, Dr Tosso Leeb, Dr Louise Burmeister, Dr James Oliver and Dr Oliver Forman. Parts of the text here (specifically parts of the methods and results sections) follow parts of that manuscript entirely written by Rebekkah Hitti-Malin, with some additions and modifications in the text in this thesis.

3.2 Materials and methods

3.2.1 PRA diagnosis and sample collection

PRA was initially diagnosed in a four-year-old GS male dog at the AHT by a board-certified veterinary ophthalmologist who is also a qualified BVA/KC/ISDS eye scheme panellist and a diplomate of the European College of Veterinary Ophthalmologists. Ophthalmoscopic examination detected widespread tapetal hyperreflectivity, retinal vascular attenuation and atrophy of the optic nerve head; consistent with PRA (Figure 3.1). A reduced pupillary light response was also observed. Follow-up examination of littermates by separate board-certified veterinary ophthalmologists identified two additional male dogs that were affected by PRA; meaning a total of three of the seven dogs in the litter, which comprised six males and one female, were affected by PRA. The sire and dam were clinically unaffected at eight and eleven years of age, respectively. A review of the medical records of each dog showed no history of neurological disease in any dog. gDNA was extracted from buccal mucosal swabs following the protocols described previously (see section 2.2).

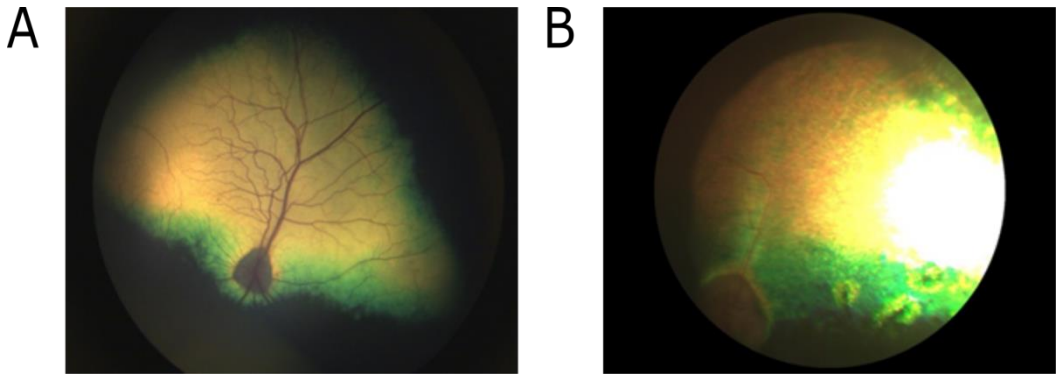


Figure 3.1: Fundus changes observed in GS dogs showing **(A)** non-affected GS dam at 11 years; **(B)** GS PRA case at five years of age. Observations of PRA cases include bilateral tapetal hyperreflectivity and retinal vascular attenuation.

For the Finnish cohort used for subsequent variant screening, eye examinations were conducted by veterinary ophthalmologists board-certified by the European College of Veterinary Ophthalmologists; 3 mL EDTA-blood was collected under the permission of animal ethical committee of County Administrative Board of Southern Finland (ESAVI/343/04.10.07/2016) and DNA was extracted from white blood cells using a semi-automated Chemigen extraction robot (PerkinElmer Cheagen Technologie GmbH, Baeswieler, Germany) according to manufacturer's instructions.

3.2.2 Exclusion of known retinal mutations

Prior to WGS, 25 previously reported variants associated with retinal disease in the dog were excluded as causal in two of the three GS dogs diagnosed with PRA by a veterinary

ophthalmologist and/or a panellist of the BVA/KC/ISDS Eye Scheme. Full materials and methods are described in Chapter 2 (section 2.7).

3.2.3 Whole genome sequencing (WGS)

DNA from four GS (two PRA cases and both non-affected parents) was prepared for WGS. WGS was performed and data were processed and analysed following the protocols described in section 2.6.

3.2.4 Variant filtering

WGS data of two PRA-affected GS full-siblings were compared with corresponding data of their non-affected parents and control dogs. Variants were filtered by the predicted effect of each variant on the transcript and protein and then based on those segregating appropriately for:

1. An X-linked recessive condition i.e. dam is heterozygous, both cases are hemizygous (appearing as homozygous on the X chromosome when visualising sequencing reads in IGV) and the sire is hemizygous for the alternate allele.
2. An autosomal recessive mode of inheritance i.e. variants were retained if cases were homozygous, parents heterozygous, and controls homozygous for the alternate allele. The DBVDC was used for additional variant filtering to screen more canine genomes from multiple dog breeds.

To exclude known canine XLPRA mutations that could not be excluded prior to WGS, sequencing reads from the two GS PRA genomes were visualised in IGV and a *de novo* assembly of sequencing reads across the mutation regions was conducted, as described in Chapter 2 (section 2.6). In addition, the genome data of six additional genes associated with X-linked retinal diseases in humans: *RP2* (Bhattacharya et al, 1984; Hardcastle et al, 1999), *OFD1* (Coene et al, 2009; Hardcastle et al, 2000; Webb et al, 2012), *TIMM8A* (Jin et al, 1996), *CACNA1F* (Bech-Hansen et al, 1998; Strom et al, 1998) and *NYX* (Bech-Hansen et al, 2000; Gal et al, 1989; Hardcastle et al, 1997) were visualised for variants that were heterozygous in the dam, hemizygous (appearing homozygous on the X allele in IGV) in both cases and hemizygous for the wild type/common allele in the sire.

3.2.5 Variant screening

Primers and allelic discrimination probes were designed as described in Chapter 2 (section 2.3). Individual custom probe-based assays were resuspended in ddH_2O to a 40X mix and combined. Allelic discrimination assays were carried out as described in Chapter 2 (section 2.9) using qPCR primers and probes and conditions listed in Appendix 4. Results were analysed using Applied Biosystems StepOne Software v2.3.

For the Finnish cohort, DNA was amplified by PCR using components listed in Table 3.1 and primers and conditions listed in Appendix 4. PCR products were sequenced by Sanger sequencing as explained in Chapter 2 (section 2.5).

Table 3.1: PCR components used to amplify the *NECAP1* variant in the Finnish cohort.

Component	Volume/reaction (µL)	Final concentration
10X PCR Buffer	1.20	1X
$MgCl_2$ (50mM)	0.40	0.008 mM
dNTP (10mM)	0.24	0.024 mM
Forward primer (5 mM)	1.90	0.380 mM
Reverse primer (5 mM)	1.90	0.380 mM
Betaine (5M)	2.00	0.400 mM
Biotools DNA Polymerase (5 units/ µL)	0.10	0.020
ddH_2O	3.06	-
Total (µL)	10.8	0.120

3.2.6 *In silico* protein prediction tools

To assess the impact of the candidate variant amino acid substitution on gene protein function, bioinformatics tools Mutation Analyzer, PolyPhen-2, and SIFT were used to computationally predict the effect of the missense variant.

3.2.7 Comparative species conservation

The Evolutionary Conserved Region Browser (http://ecrbrowser.dcode.org) and UCSC Genome Browser were used to determine the degree of sequence conservation of the *NECAP1* gene across species. Using the human genome (GRCh37/hg19) as the base genome, *NECAP1* human coordinates were compared across five species.

3.2.8 Screening cohorts

To increase the number of individual dogs that were ultimately screened, collaborators at the University of Bern, Switzerland (*n* = 199); University of Helsinki, Finland (*n* = 487) and Mars Veterinary, UK (*n* = 3,162), kindly genotyped DNA from dogs of multiple different breeds, including mixed breed dogs and dogs of unknown breeds, for the candidate variant. Breed-specific

allele frequencies for the candidate variant were determined using cohorts of dogs from the AHT, UK; University of Bern, Switzerland and University of Helsinki, Finland because the dogs comprising these cohorts were all 'purebred' dogs of known breeds. The dataset from Mars Veterinary, UK, comprised mixed-breed individuals originally submitted for breed testing on the Wisdom Panel platform, a commercially available test used to estimate breed ancestry of mixed breed dogs, and these genotyping results did not therefore contribute towards overall breed frequency calculations.

3.2.9 Sanger sequencing of *NECAP1* variant

The candidate SNV was confirmed in all three GS PRA cases by PCR amplification; primers and conditions are listed in Appendix 4. PCR products were sequenced using the Sanger sequencing method described in Chapter 2 (section 2.5).

3.2.10 Expression of *NECAP1* in canine retina

RNA sequencing (RNA-seq) data generated from retinal RNA of a PRA non-affected Petit Basset Griffon Vendéen control dog for a previous unrelated project (Forman et al, 2015) were used to determine gene expression. Using RNA-seq expression data published in GeneCards (Stelzer et al, 2016), three genes known to be implicated in retinal degeneration and considered to be expressed in retinal tissue (*RPGRIP1*, *RHO*, and *GNAT1*) and three genes considered to have low/no expression in human retinal tissue (*ANXA1*, *MYH41*, *MYOT*) were selected as control genes to evaluate expression in the control retinal RNA-seq data from a Petit Basset Griffon Vendéen dog. The total number of reads aligned to each of the six control genes was obtained using BEDTools v2.17.0 'coverage' tool (Quinlan & Hall, 2010) and compared to that of *NECAP1*. An average number of reads per exon was calculated for each gene.

3.2.11 Haplotype analyses

A total of eight SNVs, four either side of the *NECAP1* variant, that were homozygous in the cases and heterozygous in the parents were selected for genotyping. SNVs were situated up to 573 Kb downstream of the variant and 301 Kb upstream. Primers were designed using Primer3 and products amplified using PCR conditions, purification, and Sanger sequencing, as described in Chapter 2 (sections 2.3, 2.4 and 2.5), and primers listed in Appendix 4. SNVs surrounding the variant were genotyped across two GS PRA cases, both GS parents, one control GS, one heterozygous Giant Spitz, six heterozygous Miniature Long-Haired Dachshunds, and a control Miniature Long-Haired Dachshund to determine a disease-associated haplotype.

3.2.12 Autozygosity mapping

Illumina Canine HD array SNP positions were extracted from the four GS WGS datasets into binary files using PLINK2 (www.cog-genomics.org/plink/2.0/) (Chang et al, 2015) and an in-house PERL script. PLINK v1.07 (Purcell et al, 2007) was used to identify runs of homozygosity (ROH) using

sliding windows across the SNP data using the default values: -homozyg-window-kb 5000 (length in Kb of the sliding window); -homozyg-window-snp 50 (minimum number of SNPs that the sliding window requires); -homozyg-window-het 1 (allowing 1 heterozygous SNP in a window to tolerate genotyping errors); -homozyg-window-missing 5 (allowing up to 5 missing calls in a window); -homozyg-window-threshold 0.05 (proportion of overlapping windows required as homozygous to define a given SNP in a ROH); -homozyg-snp 100 (sliding window of 100 SNPs); -homozyg-kb 1000 (sliding window of 1,000 Kb/ 1 Mb); -homozyg-density 50 (minimum density of 1 SNP per 50 Kb); and -homozyg-gap 1000 (split a segment into two if the distance between two SNPs is > 1,000 Kb, i.e., retaining ROH >1 Mb in size). ROH overlapping with one or both parents were removed therefore leaving regions that were homozygous in the cases and heterozygous in the parents. Genes within each ROH block were cross-referenced with the RetNet database (RetNet, 2020) and RetNet gene regions were visualised in the GS WGS data for high impact SNVs or structural variants.

3.3 Results

3.3.1 Variant identification

Screening for 26 previously published retinal mutations (Chapter 2, section 2.7) revealed all PRA-affected GS were clear of all these mutations. At the time of study, 118 canine genomes sequenced as part of other studies at the AHT were available to use as non-breed matched controls to compare against the PRA-affected GS genomes. Of these 118 genomes, 112 were of dogs that were recruited for non-PRA studies, considered to be PRA controls for the purposes of this study, comprising 110 dogs across 58 different dog breeds and 2 cross breed dogs. Date of births were known for 69 of the 112 non-breed matched controls, with dogs ages ranging from 0.1 - 15.4 years at the time of DNA submission, with a median age of 6.3 years (interquartile range 1.6 - 8.7 years). These dogs were affected with conditions including epilepsy, sensory neuropathy, hereditary cataracts, glaucoma and primary lens luxation. WGS data from two PRA-affected full-siblings were compared with corresponding data of their non-affected parents and 112 control dogs (a total of 116 genomes/VCFs in analysis). A total of 25,658,899 variants were detected amongst 116 canine genomes of 58 breeds and 2 cross breeds, including the four GS genomes. The DBVDC was used for further filtering that, at the time of study, contained 452 WGS, 13 of which were contributed by the AHT.

X-linked recessive model

As the three PRA-affected GS dogs in the litter were males, and both parents were clinically clear of PRA, an X-linked recessive mode of inheritance was considered. Following WGS, previous XLPRA mutations (XLPRA1 and XLPRA2; Zhang et al, 2002) were excluded by conducting a *de novo* assembly of sequencing reads across the mutation regions in the *RPGR* gene. As well as manually interrogating the entire *RPGR* gene in the GS genomes by visualising sequencing reads in IGV software, the genome data of six additional genes associated with X-linked retinal diseases in humans: *RP2*, *OFD1*, *TIMM8A*, *CACNA1F* and *NYX* were visualised for variants that were heterozygous in the dam, hemizygous (appearing homozygous on the X allele in IGV) in both cases and hemizygous for the wild type/common allele in the sire. No variants segregated in this way. WGS analysis filtered for variants predicted to have a high/severe effect on the protein-coding sequence and present only on the X chromosome. One variant that had the correct segregation pattern for an X-linked recessive disorder also had a high effect score (in-frame insertion): a 9-bp insertion (CAGCAGCAG) in exon one of the androgen receptor (*AR*) gene. Visualisation of this insertion in the WGS data in IGV showed this to be within the second of two polymorphic tri-nucleotide CAG repeats, within exon 1 of the canine *AR* gene (CANFAX: 51,970,321). Polymorphic CAG repeat length variations in *AR* cause inherited disease in humans, including associations with spinal and bulbar muscular atrophy (SBMA) (La Spada et al, 1991), however in humans there is only one CAG repeat in a location that is orthologous to the first CAG repeat in the canine genome

(Lai et al, 2008). The insertion of three additional CAG repeats in the GS genomes results in an increase from 11 to 14 CAG repeats, with previous reports of CAG repeats at this location reported in the Single Nucleotide Polymorphism Database (dbSNP) (rs853062526). Expression of *AR* in canine retinal tissue was assessed in RNA-seq data from a Petit Basset Griffon Vendéen dog and showed no sequencing reads aligning to the gene, suggesting *AR* is not expressed in the normal canine retina.

Autosomal recessive model

As both parents were clinically clear of PRA, and no candidate variants segregating with the disease were identified through an X-linked recessive disease model, an autosomal recessive model was implemented. Variants predicted to be highly penetrant were prioritised, however employing this filtering method has limitations. These include the lack of further prioritisation of non-coding variants in known IRD/PRA genes and the possibility of dismissing structural variants or non-coding variants in potential novel IRD/PRA genes. Filtering for protein changing variants on all chromosomes reduced the initial 25,658,899 variants to 102,716 variants that were predicted to have a high/severe effect on coding sequence and alter the protein. These included variants that introduced a loss/gain of the start/stop codon, missense variants, in-frame INDELs and splice-site variants. Of these, 12 variants had an effect score of 3 and above, were homozygous in both cases, heterozygous in both parents and either homozygous for the common allele (n = 5) or heterozygous (n = 7) in the initial control set of 112 canine genomes.

Genomes in the DBVDC, comprising 452 genomes of 94 individual breeds, eight cross breeds, and three wolves were used for additional filtering to highlight common variants. Exclusion criteria included the presence of dogs that were homozygous for a specific variant and/or more than 10 heterozygous dogs (that were not GS dogs) identified within the consortium genome bank. Genotypes of the initial 12 variants were also extracted from genomes in the DBVDC. This resulted in the exclusion of seven variants, which had already been previously excluded using the initial control set 112 canine genomes due to the presence of multiple heterozygous individuals of other breeds. Details of the seven variants that were excluded are listed in **Table 3.2**.

Table 3.2: Seven variants excluded from filtering analysis. Further filtering using the DBVDC dataset excluded seven of the twelve variants identified through variant filtering based on the presence of alleles in additional dogs.

CanFam3.1 Chromosomal Coordinates	Gene	Variant	Total number of homozygous canids in DBVDC	Total number of heterozygotes in AHT and DBVDC genomes
23: 5,421,356	ENSCAFG00000010012	**Missense variant**	**Entlebucher Sennenhund (*n*=1) and Miniature Schnauzer (*n*=1)**	**39**
23: 5,421,500	ENSCAFG00000010012	**Missense variant**	**Miniature Schnauzer (*n*=1)**	**40**
27: 40,341,520	*RAD51AP1*	**Missense variant**	**Yorkshire Terriers (*n*=3)**	**26**
30: 7,520,689	*C15orf52*	**Missense variant**	**Lagotto Romagnolo (*n*=1), Norwich Terrier (*n*=1), Bichon Frise (*n=1),* West Highland White Terrier (*n*=2),**	**23**
8: 62,125,521	*CCDC88C*	**Missense variant**	**Bearded Collie (*n*=1), Golden Retriever (*n*=1), Newfoundland (*n*=1), West Highland White Terrier (*n*=1), Yorkshire Terrier (*n*=2)**	**47**
23: 9,908,657	*RPL14*	**Splice site variant**	**0**	**24**
23: 5,421,400	ENSCAFG00000010012	**3-bp insertion**	**Miniature Schnauzer (*n*=1), Yorkshire Terrier (*n*=1)**	**39**

The GATK HaplotypeCaller tool was used to perform a local reassembly and realignment of sequencing reads in the GS cases to confirm the five variants of potential interest (Table 3.3). At the time of analysis, three variants were situated within a novel gene, ENSCAFG00000030684, for which no human homologues existed nor did any annotated genes lie nearby. Since the update of a more recent annotation of the CanFam3.1 canine assembly in the Ensembl database (database version 100.31), this ENSCAF gene is no longer present in Ensembl, however the addition of a nearby gene is present: the *HNRNP3* gene located on chromosome 36 between coordinates 20,974,648-20,980,137. The locations of the three variants identified in the former ENSCAFG00000030684 gene were visualised in the updated Ensembl CanFam3.1 canine assembly, however none of these variants are located in any newly annotated genes and therefore would not have passed the original filtering. HaplotypeCaller local realignments showed both cases and parents were in fact heterozygous for two of the variants within the novel gene, as well as a variant within the *LURAP1L* gene. This left a single candidate missense variant in the gene *NECAP1* encoding the NECAP endocytosis associated 1 protein for which, aside from the two GS parents, only one additional heterozygous dog was identified (Giant Spitz breed) in the full dataset of 568 genomes (WGS controls and DBVDC cohort).

Table 3.3: Variants retained following filtration steps. Variant filtering revealed five variants from WGS analysis that were not homozygous in a control set 568 genomes. HaplotypeCaller enabled local realignment of sequencing reads to confirm/exclude each variant call from the WGS analysis.

CanFam3.1 Chromosomal Coordinates	Gene	GATK analysis consequence in GS cases	HaplotypeCaller realignment consequence in GS cases	Homozygote in DBVDC
11: 33,389,711	*LURAP1L*	Homozygous missense variant followed by a frameshift variant	Heterozygous 8-bp insertion (CAGCAGCA) followed by heterozygous T/A SNV in a repetitive region	Unable to determine
27: 37,468,611	*NECAP1*	Homozygous missense variant	Homozygous missense variant	None
36: 20,974,031	ENSCAFG00000030684	Homozygous missense variant	Homozygous missense variant	Chinese Indigenous dog (*n* = 1)
36: 20,974,357	ENSCAFG00000030684	Homozygous missense variant	Heterozygous missense variant	None
36: 20,974,370	ENSCAFG00000030684	Homozygous missense variant	Heterozygous missense variant	None

The c.544G>A non-synonymous variant is in exon 6 of 8 in the *NECAP1* gene at the following coordinates: CANFA27:37,468,611 (CanFam 3.1 assembly; Figure 3.2), and is predicted to cause a glycine (G) to arginine (R) amino acid substitution at amino acid position 182 (p.G182R).

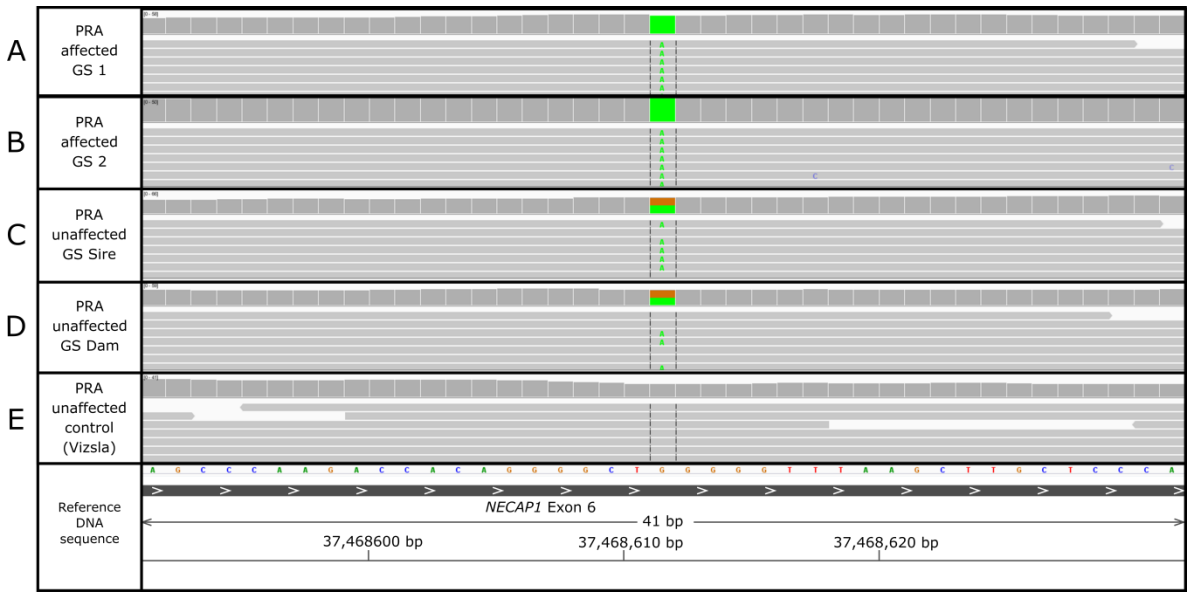


Figure 3.2: View of the *NECAP1* SNV in WGS data in IGV. Two GS PRA cases are homozygous **(A,B)** for the c.544G>A (p.Gly182Arg) SNV in exon 6 of 8 of *NECAP1*, the PRA non-affected sire **(C)** and dam **(D)** are heterozygous for the variant, and Vizsla control is homozygous for the wild type allele **(E)**.

Sanger sequencing confirmed all three GS PRA cases were homozygous for the mutant allele (A/A), both obligate carrier parents were heterozygous (G/A) and the remaining four PRA non-affected littermates were either heterozygous or homozygous for the wild type allele (G/G) (Figure 3.3, Figure 3.4).

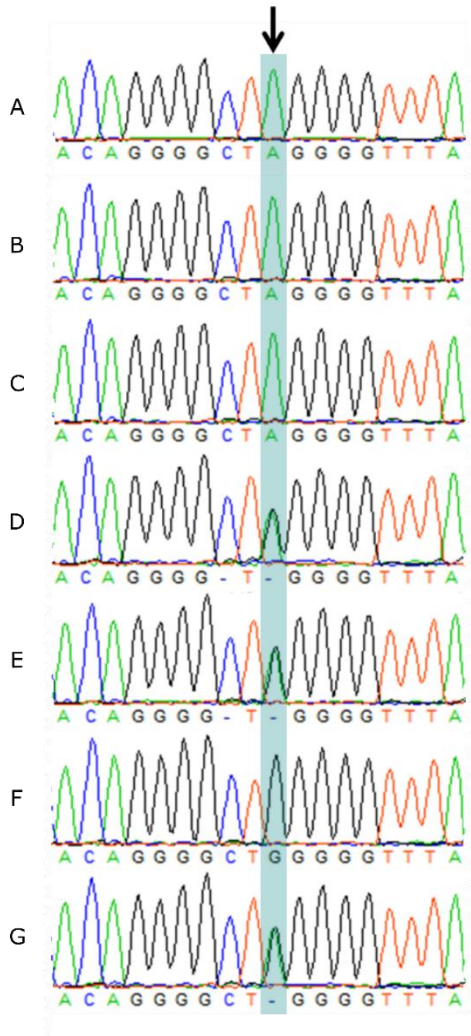


Figure 3.3: Sanger sequencing traces of the *NECAP1* c.544G>A SNV in *NECAP1* in three GS PRA cases (A, B, C); the heterozygous non-affected sire (D) and dam (E); a homozygous wild type non-affected sibling (F), and a heterozygous non-affected sibling (G).

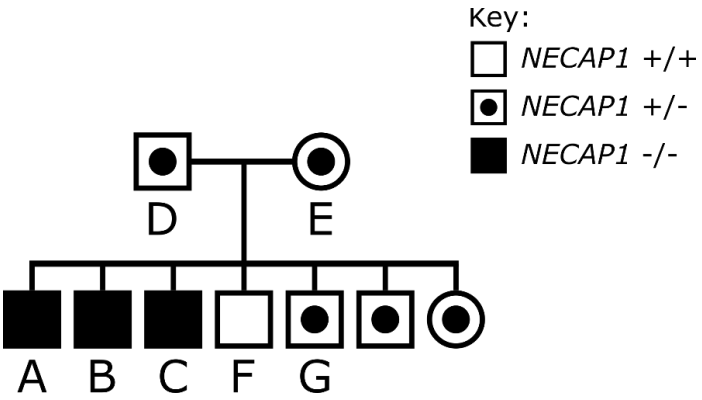


Figure 3.4: Pedigree of the GS family showing *NECAP1* genotypes. Male dogs are represented as a square symbol and female dogs as a circle symbol. PRA cases were homozygous for the *NECAP1* variant (*NECAP1*-/-); clinically clear relations were either heterozygous (*NECAP1* +/-) or homozygous for the wild type allele (*NECAP1* +/+). Labels A-G correspond to Sanger sequencing traces in Figure 3.3 as follows: three *NECAP1*-/- PRA cases (A, B, C); the *NECAP1* +/- non-affected sire (D) and dam (E); a *NECAP1* +/+ non-affected sibling (F), and a *NECAP1* +/- non-affected sibling (G).

3.3.2 *In silico* protein predictions

To assess whether the c.544G>A; p.G182R amino acid substitution affects protein function, bioinformatics tools Mutation Analyzer, PolyPhen-2, Mutation Taster2, and SIFT were used to predict the effect of the missense variant. Mutation Analyzer and PolyPhen-2 predicted the amino acid change to be rarely tolerated and probably damaging to the protein, respectively. Mutation Taster predicted the amino acid substitution to be disease-causing and SIFT predicted the change to be deleterious with a SIFT score of 0.03.

3.3.3 Comparative species conservation

NECAP1 is conserved across species showing > 80% sequence similarity to the human orthologue. The glycine amino acid at amino acid position 182 that is substituted to an arginine in the present study is highly conserved across 31 mammals, including humans, mice and rhesus macaques. In addition, the GGG codon implicated is well conserved across 37 mammals, including humans. Using the human genome (GRCh37/hg19 assembly) as the base genome, the degree of evolutionary sequence conservation of *NECAP1* was assessed in the zebrafish, chicken, mouse, dog (CanFam2.0 assembly) and rhesus macaque (Figure 3.5A). All coding exons were conserved across the human, dog and rhesus macaque, sharing the highest degree of *NECAP1* sequence conservation across the species evaluated. The mouse genome only shared sequence conservation across exons 5 and 6 of *NECAP1* and the zebrafish and chicken genomes displayed an absence of conservation across all exons. There is variability in the conservation of NECAP1 amino acid residues across exon 6, however the affected amino acid glycine is highly conserved through evolution (Figure 3.5B).

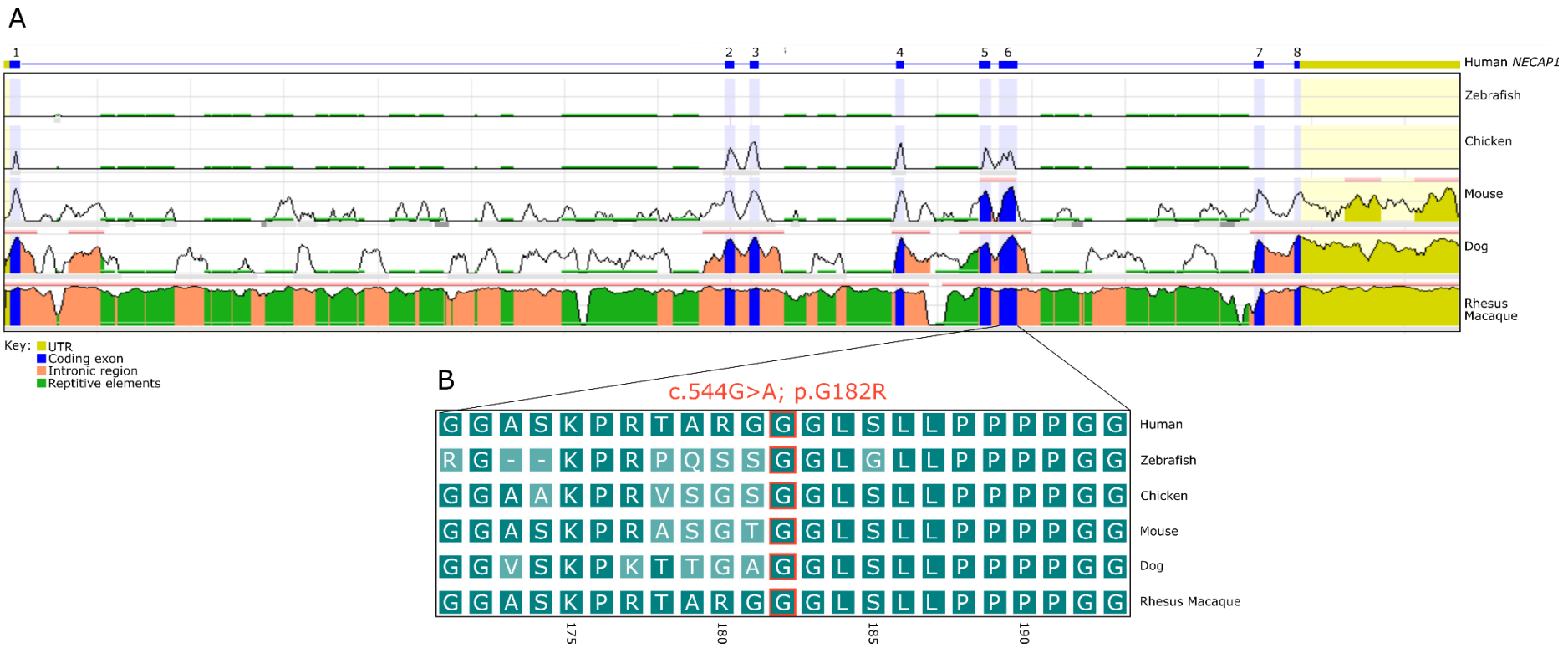


Figure 3.5: Display of *NECAP1* conservation across species. **(A)** Evolutionary conserved regions of human *NECAP1* (GRCh37; chr12: 8234807-8250373) are highly conserved across the rhesus macaque and dog (CanFam2 assembly) genomes. The mouse construct shows poorer conservation, with only exons 5 and 6 conserved. **(B)** Evolutionary conservation of the amino acid affected by the canine *NECAP1* missense variant (p.G182R; glycine boxed in red) and surrounding amino acids across the human, zebrafish, chicken, mouse, dog and rhesus macaque show a high degree of conservation. Amino acids numbered are relative to the dog sequence (CanFam2 assembly).

3.3.4 *NECAP1* c.544G>A screening

To investigate whether the variant is common in the breed, additional GS dogs were genotyped for the c.544G>A variant. Six additional GS heterozygotes were initially found. Screening an additional 427 dogs of 122 breeds, including 65 Miniature Schnauzers, identified one Miniature Long-Haired Dachshund that was heterozygous for the c.544G>A variant.

3.3.5 Population screening

The presence of the *NECAP1* variant in GS dogs and dogs of additional breeds led to the screening of additional GS dogs, Miniature Long-Haired Dachshunds and varieties of German Spitz dogs. In total, the c.544G>A candidate variant was genotyped in 5,130 canids. Where known, 1,974 canids were of 173 breeds, 10 cross breed dogs, and 3 wolves, including a total of 323 GS dogs (Appendix 7).

Of the GS screened, only the three PRA-affected GS were homozygous for the variant and 19 heterozygotes were identified, nine of which were closely related (within three generations) to the two WGS cases, and ten were either known to be not closely related to the affected dogs based on pedigree information, or deemed unknown due to lack of pedigree information. Allele frequency in GS dogs, excluding closely related dogs, was 0.015.

Genotyping cohorts of additional breeds of German origin, including Spitz and Dachshund varieties, identified dogs that were heterozygous for the variant. Genotyping data were compiled, and allele frequencies determined (Table 3.4). Four PRA affected dogs (two Miniature (Klein) Spitz and two Pomeranian Spitz) were included in the testing cohorts. These were homozygous for the wild type allele, suggesting their PRA phenotype is due to a different PRA-associated variant. Out of the total 5,130 canids genotyped for the *NECAP1* variant, the only dogs homozygous for the variant were the three GS PRA cases.

Table 3.4: Total allele frequency of *NECAP1* SNV in multiple breeds of dog.

Breed	*NECAP1* +/+	*NECAP1* +/−	*NECAP1* −/−	Total	Allele Frequency
Giant Schnauzer (GS)	301	19	3	323	0.039
GS *	300	9	0	309	0.015
Giant Spitz	109	7	0	116	0.030
Medium Spitz	146	5	0	151	0.017
Miniature Spitz	109	8	0	117	0.034
Pomeranian Spitz	56	3	0	59	0.025
Miniature Long-Haired Dachshund	160	8	0	168	0.024
172 other breeds, mixed breeds and unconfirmed breeds	4,181	15	0	4,196	-

* excluding third generation relatives to proband from known pedigree information.

One of the study cohorts (Mars Veterinary, UK) consisted of a combination of unconfirmed breeds and mixed breed dogs (n = 3,162) in which 21 additional heterozygotes were identified. Six heterozygotes were purebreds, comprising Miniature Long-Haired Dachshund (n = 2), Standard Smooth-Haired Dachshund (n = 1), Miniature Spitz (n = 2), and GS (n = 1); 13 were mixed breed dogs for which breed compositions could be estimated (Table 3.5); and breed type could not be estimated for two heterozygotes. As this cohort included various mixed breed or unknown breeds of dogs, these were excluded from the breed specific allele frequency calculations.

Table 3.5: Known breed compositions of mixed breed dogs tested by Mars Veterinary, UK. Breeds are known for 13/15 *NECAP1* heterozygous mixed breed dogs.

Dog	Breed composition
1	50% Dachshund mix
2	50% Pekingese mix, (trace German Spitz)
3	25 % Poodle, 12.5% Russell Terrier, 12.5 % Vizsla mix (trace German Spitz)
4	25% Australian Kelpie, 25% Labrador Retriever mix (trace Dachshund)
5	62.5% Dachshund, 12.5% Brittany Spaniel mix
6	37.5% American Staffordshire Terrier, 12.5% Labrador Retriever, 25% Dutch Shepherd Dog mix
7	50% German Shepherd dog mix (trace dachshund)
8	50% Labrador Retriever mix
9	75% Pomeranian mix
10	12.5% Rottweiler, 12.5 % Tibetan Mastiff mix
11	25% Poodle, 25% Kritikos Lagonikos mix (trace Japanese Spitz)
12	12.5% Dachshund, 12.5% Pumi, 12.5 % Pekingese mix
13	12.5 % Dachshund, 12.5% Rhodesian Ridgeback, 12.5% Kritikos Lagonikos, 12.5% Poodle mix

3.3.6 *NECAP1* is expressed in canine retina

A qualitative measure of expression was observed by extracting the total number of reads per exon from retinal RNA-seq data from a PRA non-affected Petit Basset Griffon Vendéen dog and calculating a mean number of reads per exon for each gene (Table 3.6). A total of 749 reads aligned over all eight exons of *NECAP1*, with 288 reads aligned specifically over exon 6 (Figure 3.6), providing evidence that *NECAP1* is expressed in canine retinal tissue to similar levels as *RPGRIP1*.

Table 3.6: Example of genes expressed in control canine retina. The total number of exonic reads in three genes considered to be expressed in human retinal tissue (*RPGRIP1*, *RHO* and *GNAT1*) and three genes considered to have low/no expression in human retina (*ANXA1*, *MYH41* and *MYOT*) were extracted and the mean number of reads per exon calculated. Comparison of these six control genes and the values generated for *NECAP1* using canine RNA-seq data showed reads aligning to *NECAP1* providing evidence of gene expression in the Petit Basset Griffon Vendéen control canine retina.

Gene	CanFam3.1 Chromosomal Coordinates	Total Number of Exonic Reads	Mean Number of Reads Per Exon
RPGRIP1	15: 18,316,887-18,387,548	1,458	58
RHO	20: 5,632,150-5,637,404	461,640	92,328
GNAT1	20: 39,129,469-39,133,156	118,800	13,200
ANXA1	1: 84,744,444-84,763,121	120	9
MYH41	5: 34,748,657-34,767,076	3	0
MYOT	11: 25,591,798-25,592,328	40	4
NECAP1	27: 37,460,607-37,506,942	749	94

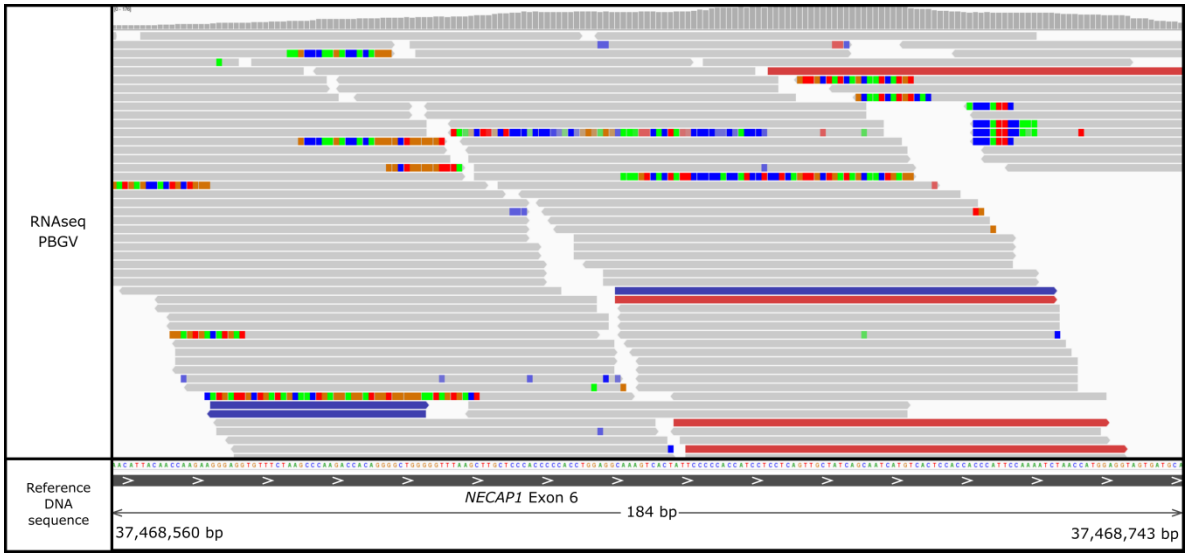


Figure 3.6: IGV display of RNA-seq data from a control canine retina (Petit Basset Griffon Vendéen (PBGV) dog) aligned to the CanFam3.1 canine genome show reads mapping to *NECAP1*.

3.3.7 Haplotype analyses

To establish whether the *NECAP1* variant identified in dogs of different breeds is due to recurrent independent mutation events or inherited identically by descent, SNVs flanking the variant were genotyped in the two GS PRA cases, both GS parents, one control GS, one heterozygous Giant Spitz, six heterozygous Miniature Long-Haired Dachshunds, and a control Miniature Long-Haired Dachshund. SNVs present as homozygous for the alternate allele in the GS PRA cases and heterozygous in the obligate carrier parents were determined. Shorter haplotypes were revealed in the Giant Spitz and Miniature Long-Haired Dachshund, with the shortest in the Dachshund, compared to that of the GS, suggesting the shared region may be an ancestral and ancient haplotype (Figure 3.7).

A

CanFam3.1 Position	Location from *NECAP1* variant	REF allele	Giant Schnauzer (GS)									
			GS Case 1		GS Case 2		GS Parent 1		GS Parent 2		GS Control	
			Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2
27:36894608	573 Kb downstream	G	A	A	A	A	G	A	G	A	G	G
27:37020314	447 Kb downstream	G	A	A	A	A	G	A	G	A	G	G
27:37425357	43 Kb downstream	T	C	C	C	C	T	C	T	C	T	T
27:37467780	4 Kb downstream	G	A	A	A	A	G	A	G	A	G	G
27:37468611	***NECAP1* VARIANT**	**G**	**A**	**A**	**A**	**A**	**G**	**A**	**G**	**A**	**G**	**G**
27:37501948	33 Kb upstream	T	C	C	C	C	T	C	T	C	T	T
27:37541149	72 Kb upstream	G	A	A	A	A	G	A	G	A	G	G
27:37644083	175 Kb upstream	A	G	G	G	G	A	G	A	G	A	G
27:37770618	301 Kb upstream	A	G	G	G	G	A	G	A	G	A	G

B

CanFam3.1 Position	Location from *NECAP1* variant	REF allele	Giant Spitz	
			Giant Spitz Control	
			Allele 1	Allele 2
27:36894608	573 Kb downstream	G	G	G
27:37020314	447 Kb downstream	G	G	A
27:37425357	43 Kb downstream	T	T	C
27:37467780	4 Kb downstream	G	G	A
27:37468611	***NECAP1* VARIANT**	**G**	**G**	**A**
27:37501948	33 Kb upstream	T	T	C
27:37541149	72 Kb upstream	G	G	A
27:37644083	175 Kb upstream	A	A	G
27:37770618	301 Kb upstream	A	G	G

C

CanFam3.1 Position	Location from *NECAP1* variant	REF allele	Miniature Long-Haired Dachshund (MLHD)													
			MLHD Control 1		MLHD Control 2		MLHD Control 3		MLHD Control 4		MLHD Control 5		MLHD Control 6		MLHD Control 7	
			Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2
27:36894608	573 Kb downstream	G	G	G	A	A	G	A	G	G	A	A	G	G	G	G
27:37020314	447 Kb downstream	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G
27:37425357	43 Kb downstream	T	T	C	T	C	T	C	T	C	T	C	T	C	T	T
27:37467780	4 Kb downstream	G	G	A	G	A	G	A	G	A	G	A	G	A	G	G
27:37468611	***NECAP1* VARIANT**	**G**	**G**	**A**	**G**	**A**	**G**	**A**	**G**	**A**	**G**	**A**	**G**	**A**	**G**	**G**
27:37501948	33 Kb upstream	T	T	C	T	C	T	C	T	C	T	C	T	C	T	T
27:37541149	72 Kb upstream	G	G	A	G	A	G	A	G	A	G	A	G	A	G	G
27:37644083	175 Kb upstream	A	A	A	A	A	-	-	A	A	A	A	A	A	A	A
27:37770618	301 Kb upstream	A	G	G	G	G	A	G	A	G	A	G	A	G	A	A

Figure 3.7: Schematic diagram of *NECAP1* alleles. To build a disease-associated haplotype, a total of eight SNVs up to 301 Kb upstream and 573 Kb downstream of the *NECAP1* variant were genotyped in **(A)** two PRA affected GS *NECAP1* homozygotes, both *NECAP1*$^{+/-}$ non-affected parents, and an non-affected *NECAP1*$^{+/+}$ GS; **(B)** one *NECAP1*$^{+/-}$ Giant Spitz from the DBVDC and **(C)** six *NECAP1*$^{+/-}$ Miniature Long-Haired Dachshunds (MLHD) and one *NECAP1*$^{+/+}$ Miniature Long-Haired Dachshund. A disease associated haplotype (shaded orange) was determined. The reference allele is shaded yellow.

3.3.8 Autozygosity mapping

Concurrently with the variant filtering approach taken, an autozygosity mapping approach was also undertaken to identify ROH present in both PRA affected GS dogs. After filtering and quality control, 106,298 SNPs were tested through PLINK and identified thirteen ROH that were over 1 Mb in size and shared exclusively by both GS PRA cases (Table 3.7, Figure 3.8). As initial analysis excluded an X-linked mode of inheritance, variants identified on the X chromosome were omitted.

Table 3.7: ROH identified in GS genomes using PLINK. PLINK analysis identified 13 ROH exceeding 1 Mb in size present in both PRA affected GS dogs. The *NECAP1* SNV lies within the 9.03 Mb ROH identified on chromosome 27.

CanFam3.1 Chromosomal Coordinates	Size (Mb)
1:29,970,719-33,682,751	3.71
6:33,954,067-39,111,638	5.16
8:59,486,837-65,640,116	6.15
10:64,657,382-69,250,914	4.59
13:58,887,427-60,841,228	1.95
21:12,842,407-19,235,893	6.39
23:54,039-17,493,091	17.44
25:23,239,981-26,009,032	2.77
27:36,722,004-45,753,342	9.03
28:13,922,125-15,662,789	1.74
29:38,572,915-41,054,487	2.48
30:1,366-26,126,946	26.13
36:17,518,975-21,096,460	3.58

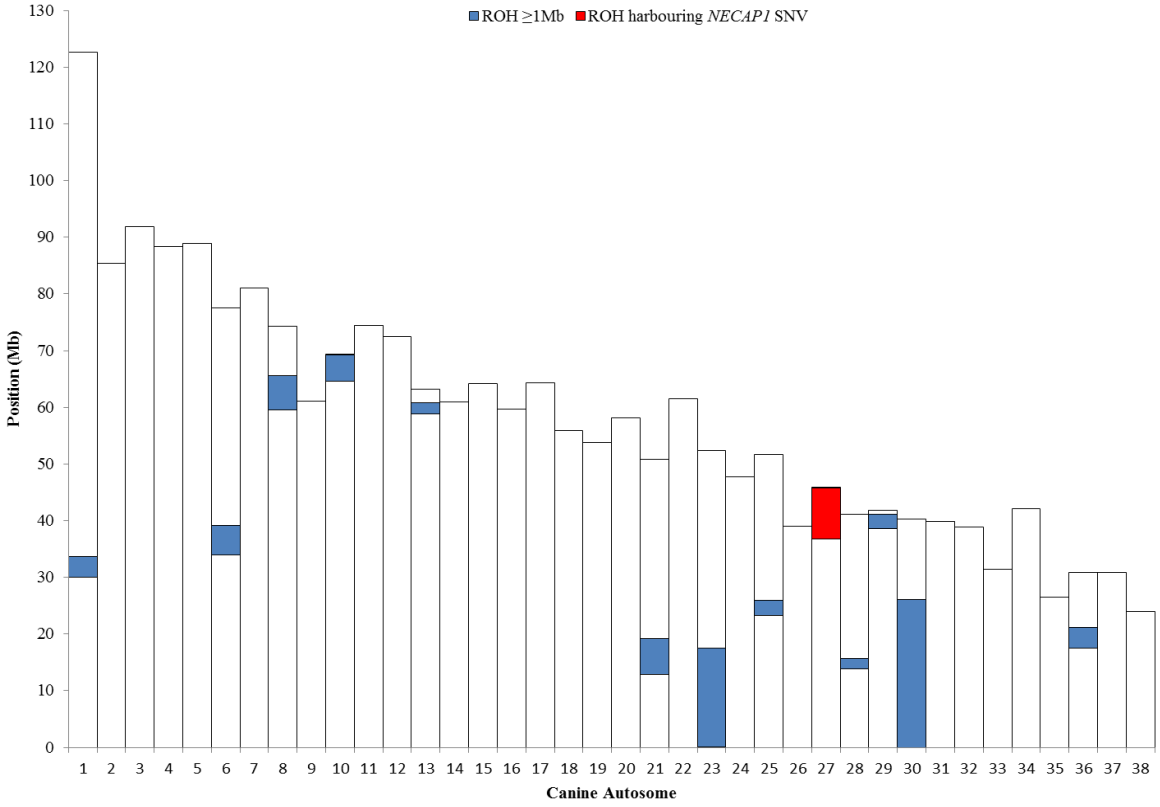


Figure 3.8: Chromosomal positions of ROH greater than 1 Mb in size (blue), on canine autosomes in the two GS PRA cases, as obtained through PLINK analysis. The red block indicates the 9 Mb ROH on CANFA27 in which the *NECAP1* variant lies.

Eight ROH harboured at least one gene present in the RetNet database (RetNet, 2020) and have been previously associated with human retinal disease. Upon visualisation of WGS data from both PRA cases and non-affected parents, no structural variants or protein-changing variants were detected. The *NECAP1* variant identified through the variant filtering process lies within a 9 Mb ROH on CANFA27 and remained the most likely variant to be disease-causing.

3.4 Discussion

PRA in GS dogs is not routinely screened for through the BVA/KC/ISDS eye scheme, as currently it is not seen frequently enough to be considered problematic in the breed. As a consequence, only three PRA cases have been identified in the breed to our knowledge; all siblings and with clinically clear parents. DNA samples from both parents, all three PRA cases with a robust PRA diagnosis, and four non-affected littermates were submitted for research.

The litter consisted of six males and one female, with the three PRA affected dogs all males. For this reason, an X-linked recessive mode of inheritance was considered. WGS analysis based on an X-linked recessive model and screening of six X-linked retinal candidate genes excluded this hypothesis. Therefore, a recessive model was adopted where both parents were considered to be obligate carriers of a recessive form of PRA. Previous published studies have identified PRA-associated variants in multiple breeds of dog; but, to date, there is no published evidence of any being present in GS dogs. However, a DNA test for PRCD-PRA mutation is commercially offered for over 70 breeds, including the GS (Mars Veterinary, UK), suggesting this mutation has been previously found in the GS breed. The three PRA cases in the present study were clear of 26 previously reported retinal mutations, including PRCD-PRA, suggesting the presence of a novel PRA-causing variant. Six previously published retinal mutations were not included in the screening to exclude known PRA-associated variants. These included three XLPRA mutations caused by variants in the *RPGR* gene (Kropatsch et al, 2016; Zhang et al, 2002) in the Siberian Husky, Samoyed, mixed breed dogs and Weimaraner dogs and a variant in the *C1orf36* gene responsible for rod cone dysplasia type 2 (rcd2) in Collie dogs (Kukekova et al, 2009) all of which are within regions of high GC content and were highly technically challenging to assay. *De novo* assembly of WGS reads in both GS cases were able to show that both cases were normal for these XLPRA variants. A number of other variants in the following genes were either not published at the time of study or were not known of at the time of exclusion screening: *SAG* (Goldstein et al, 2013a), *NPHP5* (Goldstein et al, 2013b), *PPT1* (Murgiano et al, 2018), *ABCA4* (Makelainen et al, 2019) and *HIVEP3* (Kaukonen et al, 2020). These regions were subsequently visualised in WGS data in both GS cases and were further excluded.

WGS is increasingly becoming the method of choice, when sample numbers are limited, to identify causative mutations associated with Mendelian diseases in many species. Although WGS is used for complex disease studies, this is generally alongside other methods such as GWAS. This technology has led to success in identifying mutations in canine inherited diseases using a very small number of cases and is a cost-effective approach when a GWAS or other positional approaches are unattainable due to small sample numbers. For this reason, a comprehensive WGS approach was elected for this study. Following WGS a quartet of GS dogs (two full-siblings and both parents), variants were firstly filtered for those predicted to have a severe impact on the

protein sequence by the assignment an in-house effect score. This ranks variants by their severity i.e. scored on pathogenicity with a score of 1-5; low-high severity, respectively (see Chapter 2, section 2.8). Secondly, as both parents were considered obligate carriers, the filtering process could be structured to retain variants that were heterozygous in both parents, and homozygous for the alternate allele in both PRA cases. All control genomes used in the analysis also aided in variant filtering, as they were expected to be either heterozygous for the causal variant, or homozygous for the common allele (typically the reference sequence). There is greater variation amongst the genomes of different breeds than within breeds, where more than a quarter of variation is due to differences between breeds (Parker et al, 2004). Therefore, the addition of WGS from both GS parents facilitated variant identification by removing breed specific variants that were also homozygous in the parents but absent from all other breeds and reduced the initial number of starting variants. A novel candidate missense variant in *NECAP1* was eventually identified; this gene has not formerly been implicated in retinal degeneration in any species.

The canine *NECAP1* gene on CANFA27 is comprised of eight exons spanning 46 Kb and shares 84% identity with the human *NECAP1* gene (GRCh38.p13 Ensembl Gene ID ENSG00000089818). In humans, the N-terminal region of *NECAP1* is located between amino acid residues 1-178 and contains the PH fold with ear-like function (PHear) domain (amino acid residues 1-133; **Figure 3.9**) (Ritter et al, 2007), a member of the pleckstrin homology (PH) superfamily. *NECAP1* encodes NECAP endocytosis associated 1, also termed adaptin ear binding coat associated protein 1, a subunit of the adaptor protein-2 (AP-2) heterotetrametric protein complex involved with clathrin mediated endocytosis (CME) in synapses (Ritter et al, 2003). CME is a vesicular transport event from the plasma membrane that primarily initiates the entry of material into cells using clathrin-coated vesicles (CCVs) that mature from clathrin-coated pits. In addition to clathrin, the coats on these CCVs contain a variety of adaptor proteins including AP-1, AP-2 and AP-3 (Mattera et al, 2004).

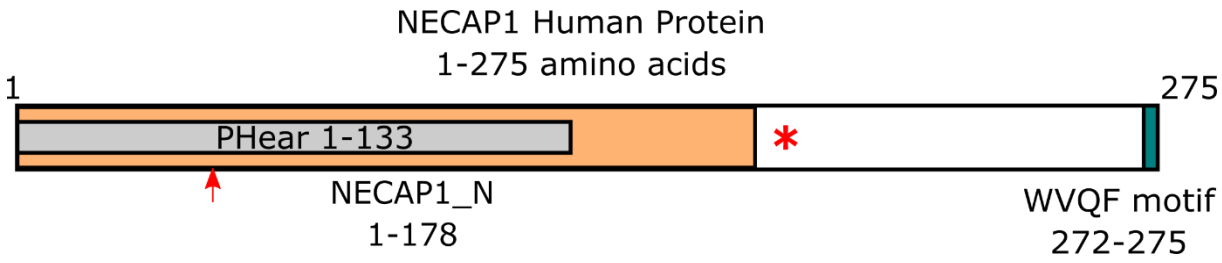


Figure 3.9: NECAP1 human protein (adapted from Ritter et al., 2013). The N-terminal (NECAP1_N) is positioned between amino acid residues 1-178, is highly conserved across species (Ritter et al., 2007) and contains the PHear domain (amino acid residues 1-133). The 'WVQF' motif is located between amino acid residues 272-275, which binds to site on the AP-2 accessory protein. The *NECAP1* candidate SNV affects amino acid position 182, as highlighted by a red asterisk. The nonsense variant in *NECAP1* described below, which has previously been associated with a recessive early infantile epileptic encephalopathy (EIEE) in humans (Alazami et al, 2014; Alsahli et al, 2018), affects amino acid residue 48, as indicated by a red arrow.

CME is the endocytic portal into eukaryotic cells for cargo molecules and facilitates nutrient uptake, signal transduction, neurotransmission, the recycling of receptors and the regulation of many plasma membrane events, as well as playing a role in synaptic vesicle reformation (McMahon & Boucrot, 2011; Ritter et al, 2004; Takei & Haucke, 2001). Complex endocytic machinery organised around AP-2 binds to the clathrin coat membrane and traps various transmembrane proteins, including cargo receptors, to enrich CCVs (Ritter et al, 2003). This triggers the formation of CCVs at the plasma membrane to engulf cargo molecules and enable the release of these into the cell following endocytosis. As clathrin does not directly bind to the plasma membrane or to cargo receptors, the ear domains of adaptor protein subunits and complexes such as AP-2 serve as a platform to allow numerous endocytic accessory proteins, cargo and clathrin to associate with the plasma membrane during CCV formation. This is facilitated via recognition of specific sequence motifs, termed 'endocytic motifs'. Endocytic motifs are specific amino acid sequences that recognise target proteins that require endocytosis and work as a signal to bind to adaptor proteins. Each endocytic accessory protein contributes to one or more specific roles in vesicle formation, including membrane deformation, cargo recruitment and vesicle size facilitated by the interaction of these adaptor proteins with domains to target vesicle formation sites (McMahon & Boucrot, 2011). One such accessory protein, NECAP1, is a CCV-enriched protein and a modulator of AP-1 and AP-2 interactions (Mattera et al, 2004; Ritter et al, 2013). The novel α-ear-binding endocytic motif 'WVQF', representing an amino acid sequence of tryptophan (W), valine (V), glutamine (Q) and phenylalanine (F) of NECAP1 (**Figure 3.9**), binds to sites on AP-2, restricting the binding of accessory proteins. When clathrin binds to vesicle formation sites, NECAP1 binding is prohibited and results in the formation of a second complex in which NECAP1 and AP-2 cooperate in accessory protein recruitment, specifically to increase the probability that a vesicle will form and to determine vesicle size and cargo content for CME (Ritter et al, 2013).

NECAP1 is significant for efficient accessory protein recruitment to vesicle formation sites to facilitate CME in normal synaptic cell function.

While *NECAP1* was initially reported to be primarily expressed in brain tissue (Ritter et al, 2003), it has also been detected throughout the central nervous system in mice, including the spinal cord (Alazami et al, 2014) and retina of mice and beagle dogs (Blake et al, 2017; Paez et al, 2006). Retinal RNA-seq data from a PRA non-affected Petit Basset Griffon Vendéen dog revealed that *NECAP1* is considered to be moderately expressed in canine retinal tissue when compared to genes known to be implicated in retinal degeneration and expressed in retinal tissue. There is also evidence that *NECAP1* (and its family member, *NECAP2*) interacts with phosphorylated AP-2 in human RPE cells where disruption of components involved in the phosphorylation event of AP-2 impacts clathrin-coated pit maturation, blocking CME (Wrobel et al, 2019). Wrobel et al. (2019) showed that phosphorylated AP-2 is present in ~70% of clathrin-coated pits in human RPE cells, which increases with clathrin-coated pit maturation, and is important for efficient CME. Previous work involving RP mouse models demonstrated that CME is present, albeit playing a minor role, in retinal ribbon synapses in photoreceptors and bipolar cells of the retina (Park et al, 2008). Park et al. (2008) showed that CME is carried out at the rod bipolar cell axon terminals in the mouse retina, suggesting that this process may be important for normal synaptic function in the mammalian retina. RP mouse models of early-onset retinal degeneration produced by a mutation in the *Pde6b* gene showed changes in the postsynaptic cells of the rod and cone photoreceptors in the retina (Park et al, 2008). Specifically, bipolar cells integrating signals of rod photoreceptor cells, involved in the rod neural pathway, generate ribbon synapses that are usually formed by postnatal day 14 in mice retinas. Preferential changes in endocytosis at the rod bipolar ribbon synapses were reported in mice exhibiting rapid photoreceptor degeneration, including abnormal synaptic ribbons in rod bipolar cells where clathrin was no longer expressed in comparison to a control mouse retina (Park et al, 2008). Moreover, an example of a gene involved in CME causing RP in humans is the receptor expression enhancing protein-6 (*REEP6*) gene (Arno et al, 2016; Veleri et al, 2017). *REEP6* is involved in intracellular trafficking of CCVs to membrane sites, and a loss of *REEP6* function resulted in photoreceptor cell death in mouse models (Veleri et al, 2017).

Studies using Drosophila as an RP disease model highlight the significance of endocytosis of rhodopsin, a light sensitive protein, which is a crucial component of the phototransduction cascade triggering vision. Endocytosis of metarhodopsin II, the active form of rhodopsin, in rhabdomeres present in Drosophila photoreceptors is essential for photoreceptor maintenance. Endocytosis of metarhodopsin is facilitated by the binding of arrestin to AP-2, hence when this pathway is compromised, metarhodopsin accumulates in the rhabdomeres and leads to the degeneration of photoreceptors (Xiong & Bellen, 2013). Examples of rhodopsin accumulation caused by mutations in genes involved in endolysosomal pathways (Chinchore et al, 2009;

Dourlen et al, 2012) suggest that reduced endocytosis in photoreceptors lead to an accumulation of rhodopsin, resulting in photoreceptor cell death and ultimately retinal degeneration.

Although there is no direct functional evidence supporting the involvement of *NECAP1* in retinal function, mutations in endolysosomal trafficking genes with similar molecular mechanisms to *NECAP1* (discussed above) suggest this is a provocative novel candidate gene for retinal degeneration. Studies of mice suggest that CME is essential for processes at the rod bipolar ribbon synapses in the mammalian retina (Park et al, 2008; Veleri et al, 2017); therefore, it can be ventured that if the expression or function of a gene regulating this process is disrupted, its role in maintaining rod bipolar ribbon synapses and subsequently photoreceptor function may be compromised. It can be speculated that the NECAP1 glycine to arginine substitution identified in GS dogs with PRA impacts protein function, and that potential inactivation of the AP-2 adaptor complex could disrupt endocytosis in retinal neurons, such as ribbon synapses in photoreceptor and bipolar cells. It is plausible that as a consequence of CME disturbance in the retina, rhodopsin accumulates in the photoreceptors, leading to cell death and retinal degeneration.

Currently, there are no reports of *NECAP1* directly implicated in retinal degeneration in any species. However, in humans, a c.142C>T; p.R48* nonsense variant in *NECAP1* has previously been associated with a recessive early infantile epileptic encephalopathy (EIEE) (Alazami et al, 2014; Alsahli et al, 2018). Epileptic encephalopathy is a heterogeneous group of neurological disorders characterised by intractable epileptic activity and is described as infantile when the age of onset is within the first few months of life. Clinical signs of EIEE include severe intractable seizures from early infancy with a progressive change in frequency and intensity. Following ERG and neuro-ophthalmological evaluations, no retinal abnormalities were detected in the patients homozygous for the *NECAP1* c.142C>T; p.R48* variant (Alazami et al, 2014), however, evidence of retinal degeneration has been reported in patients with other forms of EIEE in addition to typical EIEE clinical signs (Palmer et al, 2016; Turkdogan et al, 2017). One study described the involvement of the *ARV1* gene in EIEE, where an infant with a form of autosomal recessive EIEE was found to be homozygous for a donor splice site variant resulting in exon skipping and protein truncation with the loss of 40 amino acids. Ophthalmoscopic evaluation revealed that at six weeks of age, a minimal pupillary response to light was present leading to a diagnosis of an early-onset retinal degeneration, consistent with LCA. The patient died at one year of age (Palmer et al, 2016). A second study identified a novel 1-bp deletion in *GNB5* causing a multisystem syndrome. Patients presented with signs of EIEE, as well as retinal degeneration, cardiac abnormalities, severe neurological developmental delay, and premature sudden death (Turkdogan et al, 2017). Although these discussed forms of EIEE are caused by mutations in different genes, it can be speculated that the EIEE patients homozygous for the *NECAP1* variant may not have either survived long enough to develop retinal changes, or were not old enough at the time of study to

exhibit retinal changes, aged seven years and 3.4 years, respectively (Alazami et al, 2014; Alsahli et al, 2018). Arguably, patients reported in the aforementioned studies may have had both EIEE and retinal degeneration as a result of genetically distinct diseases. In this PRA study in GS dogs, no MRI scans or neurological examinations were performed; however no obvious neurological abnormalities were noted by the veterinary ophthalmologist(s) or by their owners. Despite the absence of any epileptic or neurological signs, this study suggests that *NECAP1* is associated with retinal degeneration and is the cause of PRA in these GS dogs. Furthermore, it can be hypothesised that the predicted premature truncation of the NECAP1 protein by the nonsense mutation described by Alazami et al. (2014) has a more extreme consequence on the protein than a missense SNV, as in the GS cases, therefore could account for the severity of the disease in humans compared with the GS dogs. Given the importance of *NECAP1* in the central nervous system, it is reasonable to hypothesise that variants in *NECAP1* affect retinal function and induce ocular phenotypes.

In the GS family studied, the *NECAP1* c.544G>A variant segregates with disease phenotype and is detected in a heterozygous state in additional breeds. Sixty-five Miniature Schnauzer dogs were genotyped in follow-up validation, in which no copies of the *NECAP1* variant were identified. This is not inconsistent with what is understood about the history of the Miniature Schnauzer and the GS, which are recognised as two distinct breeds and are not interbred. Additional genotyping of Standard Schnauzers and other breeds used to develop the GS breed may be more informative; however, sample numbers available for these breeds limited this scope. A causal mutation would be expected to be flanked by homozygous regions that were inherited with the mutation. Haplotype analysis showed that the *NECAP1* c.544G>A variant is present on a shared identical haplotype in additional breeds, implying that it arose from an ancestral founder mutation event and is inherited identically by descent rather than being identical by state. This hypothesis is supported by the retrospective autozygosity mapping, which was conducted to support WGS findings and showed the *NECAP1* variant lies within a ROH unique to the cases. This also suggests the c.544G>A SNV is more likely to be identical by descent rather than identical by state and provides additional evidence for its association with the disease. It is not uncommon for deleterious mutations to be shared across seemingly unrelated breeds, for example the PRCD-PRA mutation has been detected in various diverse dog breeds (Zangerl et al, 2006). The GS breed is believed to have originated in Germany and both those Dachshund and Spitz varieties carrying the *NECAP1* variant also appear to be of German ancestry. Shared haplotype analysis identifies only the Standard Schnauzer, Airedale Terrier, and Black Russian Terrier as ancestors to the GS, however, a study by Parker at al. (2017) examined haplotypes >232 Kb, which are likely to detect only more recent events. Parker et al. (2017) did not examine haplotypes in the Giant or Medium Spitz, and the *NECAP1* haplotype established in the Miniature Long-Haired Dachshund is only half the size described in the study. It can be hypothesised that the founding haplotype was present in

a dog of German ancestry, most likely ancestors of the Dachshund varieties, and prior to the development of these other breeds. Typically, the lengths of common shared ancestral chromosomal segments in a population are short due to the occurrence of recombination events over time, where the shorter haplotype in the Dachshund indicates increased recombination on what once was a longer haplotype. The detected haplotype may have been common in the population when the *NECAP1* variant arose and, therefore, normal copies of this haplotype were also present. The breed allele frequency of 0.015 suggests that this form of PRA in the GS is rare, resulting in insufficient cases to conduct a GWAS. Additionally, this low allele frequency provides scope for the *NECAP1* c.544G>A variant to be eliminated from the breeding GS population in a reasonable timeframe. The use of the commercial DNA test launched at the AHT in May 2019, termed 'PRA5', can prevent this form of PRA becoming widespread in the GS population.

It is important to note that other inheritance patterns, namely mosaicism and a dominant partially penetrant disorder, were not considered in the GS pedigree under investigation, however these are plausible forms of inheritance, which may result in the PRA disease phenotype. Germline mosaicism occurs in the germ line cell population, and therefore the individual (in the case of the GS family, this being both parents) can be unaffected, but its offspring affected. As three out of seven littermates were affected, a dominant PRA with partial penetrance could also have been considered, where the disease-causing variant is present in individuals with and without clinical signs and the disease displaying partial penetrance. Such forms of inheritance should be considered in future studies of PRA and other Mendelian disorders. Lastly, an additional limitation to this study is that structural variants and low-priority variants were not considered due to the supporting evidence of the *NECAP1* variant, which remained the most compelling candidate, given the caveats.

Chapter 4 PRA in the Lhasa Apso

4.1 Introduction

PRA has been reported in the Lhasa Apso (LA) dog but, until now, no genetic risk factor has been identified (Downs & Mellersh, 2014; Miyadera et al, 2009). In the present study, pedigree analysis indicated PRA has autosomal recessive inheritance in LA dogs, however the exact age of onset is unknown. The LA is currently listed on the BVA/KC/ISDS Eye Scheme and is advised to have annual eye examinations by a panellist, especially when used for breeding.

This chapter describes the use of multiple genetic technologies to explore a novel form of PRA in the LA breed. Utilisation of GWAS and WGS datasets revealed a novel variant that is strongly associated with PRA in the LA. A WES dataset that was generated from three LA PRA cases and three LA controls for a separate study in our laboratory was also interrogated, however no protein-coding mutations were found. The gene implicated is a retinal candidate gene involved in human inherited retinal disease. Variant validation facilitated the development of a DNA test, which has been utilised for over two years to reduce the prevalence of this form of PRA in the breed.

Findings from this research investigation have been submitted for publication following peer-review in the journal, *BMC Genetics*. Rebekkah Hitti-Malin performed all laboratory and *in silico* experiments, formal analysis (with statistical analysis by Dr Tom Lewis), data curation, validation and manuscript preparation. The manuscript was reviewed by all authors, with editing contributions from Dr Cathryn Mellersh, Dr David Sargan, Dr Sally Ricketts and Dr Tom Lewis.

4.2 Materials and methods

4.2.1 PRA diagnosis and sample collection

The study cohort consisted of DNA samples submitted to the AHT and to the University of Cambridge (Dr David Sargan) for research. Where known, owners of LA dogs included in the study resided in the UK, Austria, Germany and The Netherlands. All LA dogs were examined by a veterinary ophthalmologist through a clinical referral process or via the BVA/KC/ISDS Eye Scheme in the UK, or the European equivalent(s). Ophthalmoscopic examinations of PRA cases detected bilateral retinal atrophy, widespread tapetal hyperreflectivity and retinal vascular attenuation. In some cases, secondary bilateral cataracts were also observed. The age of diagnosis was known for 19 of the 21 cases and ranged from 1.75 - 11.96 years of age with a median age of 7.11 years (interquartile range 5.01 - 7.99 years). An arbitrary age of eight years or older was therefore elected for LA dogs without signs of PRA to be used as "controls" based on this. Buccal mucosal cheek swabs were used to collect gDNA from cases and controls. gDNA was extracted using protocols described in Chapter 2 (section 2.7).

4.2.2 Exclusion of known retinal mutations

Twenty-five previously reported variants associated with retinal disease in the dog were excluded as causal in two LA dogs clinically affected with PRA, as determined by a veterinary ophthalmologist and/or a panellist of the BVA/KC/ISDS Eye Scheme. Full materials and methods are described in Chapter 2 (section 2.7).

4.2.3 Genome-wide association study (GWAS)

A GWAS approach was used to investigate PRA in the LA using a case-control study of 44 LA dogs (17 PRA cases and 27 controls), as described in Chapter 2 (section 2.10).

A secondary GWAS analysis was conducted using three LA PRA cases that were genotyped as part of the initial GWAS but were not homozygous for the critical region. A new phenotype file was generated as before, indicating these three dogs as cases and the remaining control LA from the GWAS as controls. New MAP and PED files were also generated as previously described. Due to the small number of cases in this analysis, a standard case-control association analysis using a Fisher's exact test was performed using the following command replicating QC parameters from the initial analysis and performing 100,000 permutations (--mperm), where X is the file prefix:

--plinkdog --file X --fisher --geno 0.03 --maf 0.05 --mind 0.1 --mperm 100000 --out X_fisher

4.2.4 Whole genome sequencing (WGS)

Whole genome sequencing (WGS) was performed as described in Chapter 2 (section 2.6) of one LA PRA case that was included in the GWAS and was homozygous for the critical region. Sequencing reads and variants were visualised manually in IGV (Robinson et al, 2011;

Thorvaldsdottir et al, 2013) across the defined disease-associated region identified by the GWAS analysis and compared to 102 WGS non-breed matched controls.

4.2.5 Characterisation of the *IMPG2* LINE-1 insertion

The length of the long interspersed nuclear element-1 (LINE-1) insertion was estimated by PCR amplification using PrimeSTAR® GXL DNA Polymerase (Takara Bio Europe, Saint-Germain-en-Laye, France) and primers listed in Appendix 4. Reaction components in Table 4.1 were combined and PCR performed as described in Appendix 4. PCR products were separated using agarose gel electrophoresis, as described in Chapter 2 (section 2.3).

Table 4.1: Reaction components to amplify the *IMPG2* LINE-1 insertion.

Component	Volume/reaction (µL)	Final concentration
dNTPs (1.5 mM)	4.00	0.20 mM
PrimeStar GXL 5X buffer	10.00	1X
Forward Primer (20 µM)	0.75	0.30 µM
Reverse Primer (20 µM)	0.75	0.30 µM
PrimeStar GXL polymerase (1.25 units/µL)	1.00	1.25 units/µL
ddH_2O	31.50	-
Total (µL)	48.00	

Long PCR products were generated in the same way for NGS to determine the LINE-1 DNA sequence. Long PCR products were purified and prepared for NGS on a MiSeq platform (Illumina) using the methods listed in Chapter 2 (section 2.6). *De novo* assembly was performed using SOAPdenovo (Xie et al, 2014). L1Base2 (http://l1base.charite.de/) was used to analyse the LINE-1 insertion sequence in the dog genome (Penzkofer et al, 2017).

4.2.6 Variant screening

Candidate variants within the disease-associated region were genotyped in LA PRA cases and controls. Genotyping of the *IMPG2* LINE-1 insertion was performed using PCR amplification followed by AFLP analysis, as described in Chapter 2 (section 2.9). using reaction components listed in Table 4.2 and primers and conditions listed in Appendix 4.

Table 4.2: Amplification of *IMPG2* LINE-1 insertion for AFLP analysis.

Component	Volume/reaction (µL)	Final concentration
dNTPs (1.5 mM)	1.60	0.20 mM
HotStarTaq Plus 10X PCR buffer	1.20	1X
Forward Primer 1 (wild type) (10 µM)	0.11	0.09 µM
Forward Primer 2 (affected) (10 µM)	0.22	0.18 µM
Reverse Primer (10 µM)	0.22	0.18 µM
HotStarTaq Plus Polymerase (5 units/µL)	0.24	0.10 units/µL
ddH_2O	6.42	-
Total (µL)	10.00	

Allelic discrimination assays were performed as described in Chapter 2 (section 2.9) using primers and conditions listed in Appendix 4.

Genotypes for the most associated GWAS SNP (BICF2G630247609) and the *CEP97* intronic SNV (CANFA33: 8,044,097) were determined by PCR amplification using primers and conditions listed in Appendix 4. PCR products were sequenced using the Sanger method as described in Chapter 2 (section 2.5).

NNSPLICEv.0.9 (Reese et al, 1997; Berkeley Drosophila Genome Project, 2020) was used to evaluate splice site prediction to determine if intronic variants of interest caused disruption or introduction of exonic splicing or cryptic splicing

4.2.7 Predicting the *IMPG2* promoter region

To predict the promoter region of the *IMPG2* gene, DNA sequence annotated in the CanFam3.1 canine genome assembly positioned upstream of the gene was queried through *in silico* tools Gene2Promoter (Genomatix software suite, 2017) and PromoterInspector (Scherf et al, 2000). Transcription factor binding sites (TFBSs) located within the sequence were predicted using the *in silico* tool MatInspector (Cartharius et al, 2005).

4.2.8 Reporter assay

Cell culture

Canine skin fibroblast cells, Madin-Darby canine kidney (MDCK) cells and canine mesenchymal stem cells (MSC) cryopreserved in liquid nitrogen were fast-thawed over a period of 3-5 minutes. Cells stored in cryovials were removed from liquid nitrogen and left to thaw in a 37 °C water bath until a fluid phase was reached. The cell suspension was pipetted to 10 mL cell culture growth medium (see recipe in Chapter 2, section 2.12) and immediately centrifuged for 5 minutes at 160 ×g to remove the DMSO containing freezing media. The cell pellet was resuspended in cell culture growth medium, plated in a 56.7 cm^2 cell culture dish and incubated at 37 °C.

Cells were observed using a phase contrast microscope (EVOS XL Care, Life Technologies, Thermo Fisher Scientific). Once at confluence, media was removed using a pipette and cells were washed briefly using PBS (Gibco, Thermo Fisher Scientific). Cells were then detached using Trypsin with EDTA 0.5% (1X, Invitrogen, Thermo Fisher Scientific) for 3-5 minutes at room temperature or phenol red TrypLE™ Express Enzyme (1X, Invitrogen, Thermo Fisher Scientific) for up to 5 minutes at 37 °C. Detached cells were resuspended in culture media and centrifuged at 160 ×g for 5 minutes at room temperature to pellet cells. The cell pellet was resuspended in 1 mL TRI-Reagent (Sigma-Aldrich) and stored at -80 °C until further use.

RNA extraction and isolation

Harvested canine skin fibroblast cells, MDCK and MSCs stored at -80 °C in TRI-Reagent (Sigma-Aldrich) were thawed on ice and incubated at room temperature for 5 minutes. Phase separation was initiated by the addition of 300 µL chloroform (Thermo Fisher Scientific) followed by incubation at room temperature for 5 minutes and centrifugation at 12,500 ×g for 15 minutes at 4 °C. The aqueous phase was removed and combined with 1 volume of 70% ethanol (Thermo Fisher Scientific) by pipetting. The solution was transferred to a RNeasy column (Qiagen RNeasy kit, Qiagen, Manchester, UK) and centrifugation performed at 13,500 ×g for 15 seconds. The optional on-column DNase digestion was performed using the RNeasy kit (Qiagen), following manufacturer's instructions. A wash step was performed using 350 µL Buffer RW1 (Qiagen) and centrifuged for 15 seconds, 500 µL Buffer RPE and centrifuged for 15 seconds, and a final addition of 500 µL RPE centrifuged for 2 minutes, all at 16,000 ×g. RNA was eluted using 30 µL RNA free water and incubated at room temperature for 1 minute before centrifugation at 13,000 ×g for 1 minute.

Complementary DNA (cDNA) synthesis

cDNA was synthesised from RNA using the QuantiTect Reverse Transcription Kit (Qiagen) following manufacturer's instructions.

Ascertaining *IMPG2* expression

Target primers for qPCR were designed based on sequence of the experimental reaction spanning exons 16-17 of *IMPG2* cDNA, and the housekeeping gene *TBP* using primers and conditions listed in Appendix 4. qPCR was performed as described in section 2.3 using retinal cDNA synthesised from retinal tissue harvested from a control Golden Retriever dog as the control cDNA.

Construction of pGL4/*IMPG2*-promoter plasmid

To test the impact of the predicted *IMPG2* promoter sequence on gene expression, an experimental plasmid was constructed, initially using the authentic *IMPG2* predicted promoter as the insert. The pGL4 mouse interleukin-17 (mIL-17) plasmid containing a 2 Kb mIL-17 promoter with an NheI 5′ cloning site (5'-GˇCTAGC-3') and BglII 3′ cloning site (5'-AˇGATCT-3') was a gift from Warren Strober (Addgene plasmid # 20124; http://n2t.net/addgene:20124; RRID:Addgene_20124 (Zhang et al, 2008); Figure 4.1) and was used in this study as residual plasmid was available for immediate use from an unrelated study at the AHT. The pGL4 mIL-17 plasmid also contained the firefly luciferase gene (*fluc*) with a Simian virus 40 (SV40) poly-A signal, an ampicillin resistance gene (AmpR) enabling detection of plasmid-containing bacteria when grown on selective media and providing those bacteria with a means of replicating the plasmid, and an origin of replication (ori) where the DNA replication begins to allow a plasmid reproduction.

To remove the mIL-17 promoter from the pGL4 mIL-17 plasmid to generate the plasmid backbone to use as the vector, a sequential digestion of the pGL4 mIL-17 plasmid was performed initially with NheI using 10 µg plasmid, 5 µL NEBuffer2.1 (NEB), 2 µL NheI (NEB) and 33.63 µL ddH_2O and incubated at 37 °C for 8 hours. The digested product was purified using the Monarch® PCR & DNA Cleanup Kit (5 µg) (NEB) according to manufacturer's instructions and eluted in 6 µL elution buffer. Further digestion was performed with 2 µL BglII enzyme (NEB), 5 µL NheI digested vector, 5 µL NEBuffer 3.1 (NEB), and 38 µL ddH_2O and incubated at 37 °C overnight. The resulting product was separated on a 0.8% 1:1 low-melting/standard agarose gel and the DNA fragment excised under UV light using a gel extractor tip. The pGL4 backbone (vector) was dephosphorylated to prevent religation: a mix of 60 µL vector, 2 µL Antarctic phosphatase 10X buffer (NEB) and 1 µL Antarctic phosphatase (NEB) was prepared and incubated at 37 °C for 1 hour followed by an inactivation step at 80 °C for 2 minutes then held at 4 °C until further use.

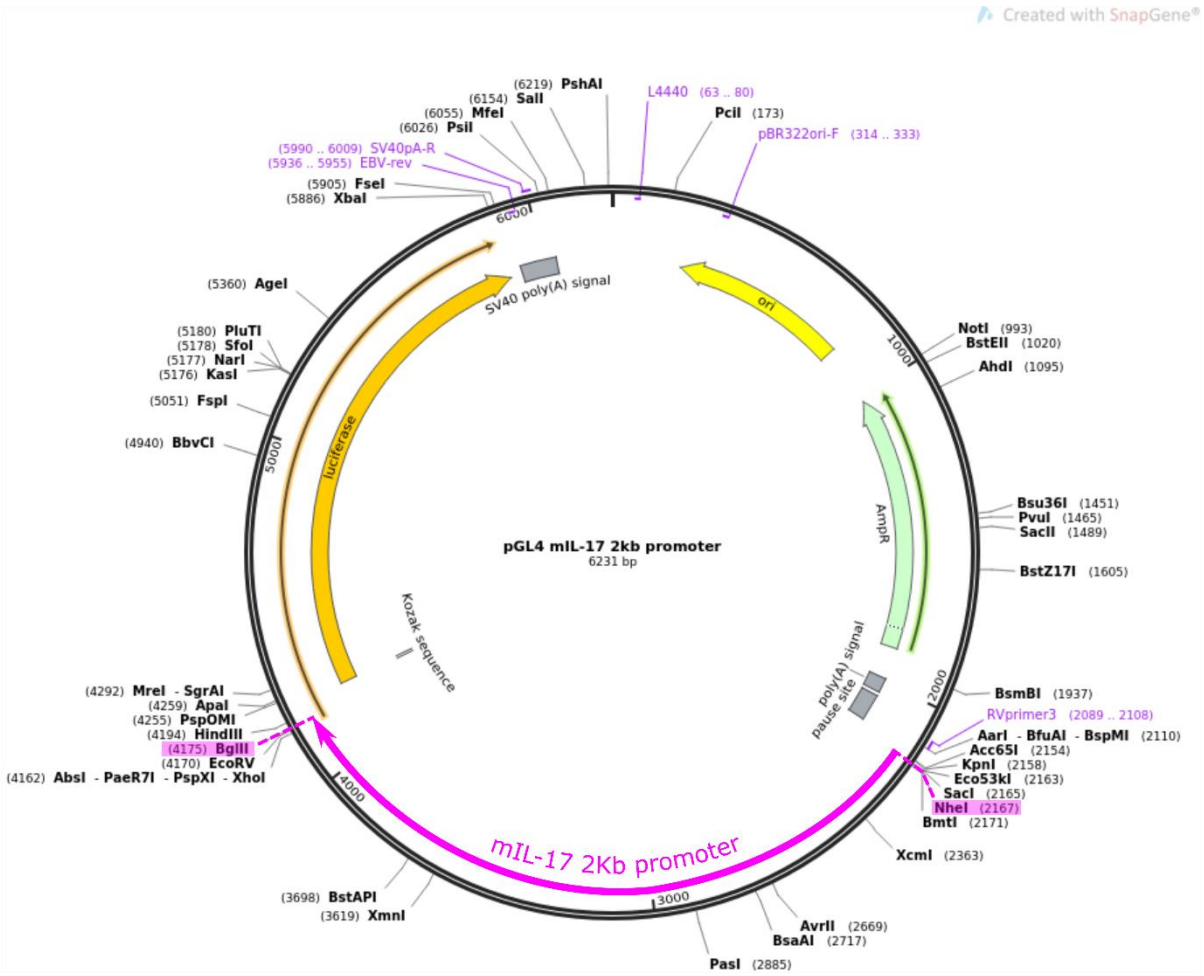


Figure 4.1: Plasmid map of the 6,231-bp pGL4 mIL-17 plasmid. The mIL-17 2 Kb promoter (labelled in pink), that would drive the luciferase gene (orange) under appropriate luciferase conditions, was removed by restriction digest at the enzyme cut sites NheI and BglII (highlighted in pink). This generated the 4,231-bp pGL4 backbone (vector) to ligate with the *IMPG2*-promoter sequence (insert). The backbone also contains the luciferase gene with an SV40 poly-A signal, an ampicillin resistance gene (AmpR) and an origin of replication (ori).

The authentic *IMPG2* predicted promoter sequence was amplified by PCR in control LA gDNA as follows. Removal of the mIL-17 promoter from the pGl4 mIL-17 plasmid resulted in NheI and BglII overhangs, however as a BglII restriction enzyme site was also present within the *IMPG2* promoter sequence, the BglII restriction enzyme could not be used to generate the BglII overhang necessary to ligate to the pGL4 vector. The BamHI restriction enzyme site shares the same overhang as BglII and did not cut any additional sites in the promoter sequence, therefore a BamHI restriction enzyme site was used as the overhang generated from the PCR *IMPG2* insert clone. A 6-bp overhang sequence (tagtag) followed by a NheI restriction enzyme site (5'-GˇCTAGC-3') was added to the 5' end of the forward primer (tagtagGCTAGCGGATTTTCAAGAGAAATATGTTTTAGATC) and the 6-bp overhang with a BamHI restriction enzyme site (5'-GˇGATCC-3') to the reverse primer (tagtagGGATCCTTGGGCCACAATCAAAGG). Each reaction contained 4 µL dNTP mix (200 µM) 10

µL 5X PrimeSTAR® GXL buffer (Takara Bio Inc.), 0.75 µL each primer (0.3 µM), 31.5 µL ddH_2O and 1 µL PrimeSTAR® GXL DNA Polymerase (Takara Bio Inc.). PCR amplification was performed as described in Appendix 4. Forty µL PCR product was purified using Monarch® PCR & DNA Cleanup Kit (5 µg) (NEB) and the purified product eluted in 20 µL elution buffer.

The purified PCR product (insert) was digested using NheI and BamHI restriction enzymes (NEB). A sequential digest was conducted using 10 µL purified insert (1 µg), 5 µL 10X NEBuffer 2.1 (1X; NEB), 1 µL NheI and 34 µL ddH_2O and incubated at 37 °C for 15 minutes. To increase the salt concentration to match that of the BamHI compatible buffer (NEBuffer 3.1), 1 µL 5M NaCl was added followed by 1 µL BamHI (NEB) and incubated at 37 °C for 15 minutes. The digestion was loaded into a 2% 1:1 low-melting temperature/ standard agarose gel. Under UV light, the digested fragments were excised from the gel and purified as described in Chapter 2 (section 2.4.2).

Ligation was carried out with 30 ng of the insert, 18 ng of the vector, 2 µL T4 DNA ligation buffer (NEB), 1 µL T4 DNA ligase (NEB) and 15 µL ddH_2O at room temperature for 1 hour. The ligation products were transformed to *E. coli* DH5α competent cells (Subcloning Efficiency DH5α Competent Cells; Thermo Fisher Scientific). Experimental reactions of 5 µL ligation reaction and 50 µL competent cells and a negative control reaction where no plasmid was added were prepared. Reactions were incubated on ice for 10 minutes, transferred to a 42 °C water bath for 90 seconds followed by a further incubation on ice for 2 minutes and 900 µL Lysogeny broth (LB) added (without antibiotic; see recipe in Chapter 2, section 2.12). A 100 µL aliquot of 950 µL suspension was cultured on LB-agar plates (recipe in Chapter 2, section 2.12) containing 100 mg/mL ampicillin. The remaining 850 µL cell suspension was centrifuged at 16,000 ×g for 10 seconds to pellet the cells and 750 µL of the supernatant was removed, leaving 100 µL to resuspend the cell pellet that was streaked across LB-agar plates containing ampicillin (100 mg/ml). This resulted in two plates per transformation to improve the chances of obtaining single colonies, while allowing all transformants to be recovered. Plates were incubated upside down at 37 °C overnight.

The following day, 15 individual colonies were picked from a single plate using the end of a sterile 10 µL tip and cultured on a single fresh LB-agar plate containing ampicillin (100 mg/ml) at 37 °C overnight. To isolate individual clonal populations, 3 of the 15 individual colonies were picked using the end of a sterile 10 µL tip and each cultured in 3 mL LB with ampicillin (100 mg/mL) pre-warmed to 37 °C in a water bath. Cultures were incubated in a shaking incubator at 37 °C at 220 revolutions per minute (rpm) overnight. Plasmid DNA was isolated from 1.5 mL of each bacterial culture using the Monarch® Plasmid Miniprep Kit (NEB), following manufacturer's instructions and eluting in 30 µL elution buffer. The purified plasmid DNA (200-500 ng) was sequenced directly by Sanger sequencing (see Chapter 2, section 2.5) using primers listed in Appendix 4 to check successful ligation of the vector and the insert and transformation of the constructed

plasmid. The remaining 1.5 mL was cultured in 500 mL LB with ampicillin (100 mg/mL) in a shaking incubator at 37 °C at 220 rpm overnight. Two of the three overnight cultures in 500 mL LB with ampicillin (100 mg/mL), that proved successful in ligation and transformation of the plasmid, based on the sequencing results, were removed from the shaking incubator the following day and transferred to plastic flasks. Cultures were centrifuged at 10,000 ×g at 4 °C for 10 minutes to pellet the cells. The supernatant was discarded, and plasmid DNA was purified from the cells using the HiSpeed Plasmid Maxi Kit (Qiagen), following manufacturer's protocol and eluting in 1 mL Buffer TE (Qiagen). A restriction digest was performed to verify correct generation of the pGL4/*IMPG2*-promoter plasmid by digesting 1 µg plasmid DNA with 5 µL 10X CutSmart buffer, 1 µL NheI and 1 µL PspOMI (all from NEB) and up to 30 µL ddH_2O at 37 °C for 2 hours. Digested plasmid DNA was evaluated on a 0.8% agarose gel against undigested plasmid to verify an expected size of 5,710-bp of the pGL4/*IMPG2*-promoter plasmid (Figure 4.2).

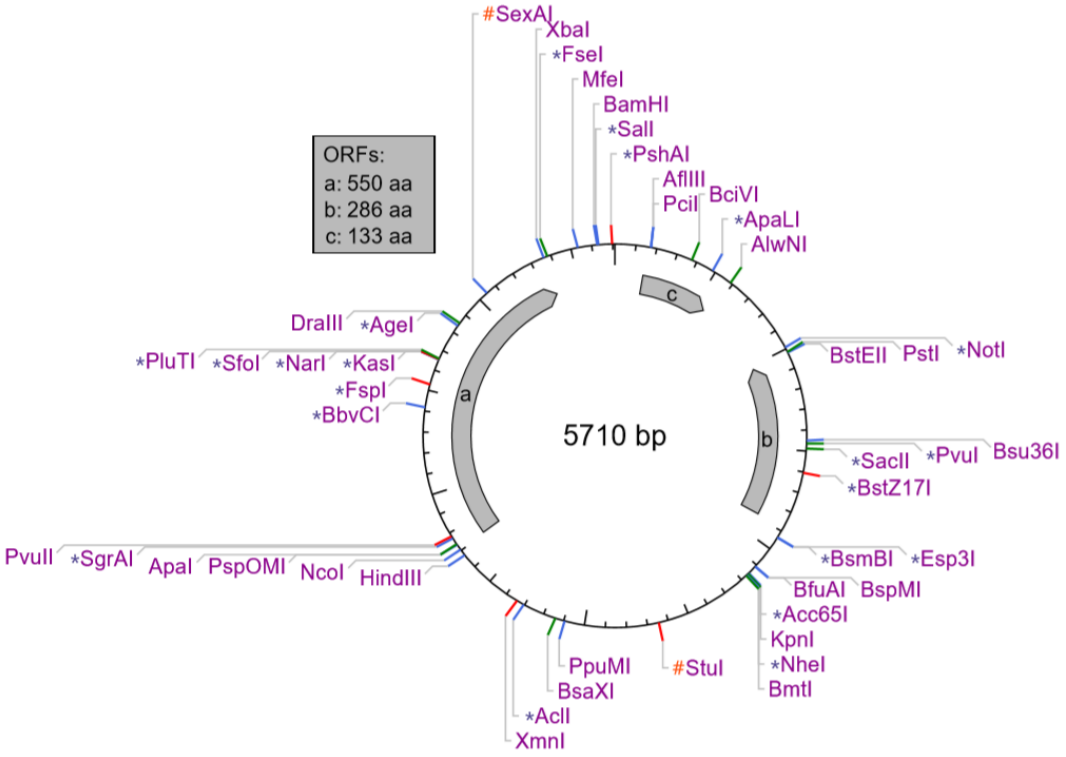


Figure 4.2: Plasmid map of the pGL4/*IMPG2*-promoter plasmid. The expected structure and size (5,710-bp) of the pGL4/*IMPG2*-promoter plasmid generated in the current study using the pGL4 backbone and *IMPG2* promoter insert. Restriction cut sites were determined using NEBCutter (NEB).

Control plasmids

Control plasmids served as experimental controls to the pGL4/*IMPG2*-promotor plasmid (experimental sample) in the luciferase reporter assay. The pGL4.10[*luc2*] basic vector (Promega UK Ltd., Southampton, UK) with no promoter that encodes the luciferase reporter gene (*luc2*) acted as an experimental 'empty vector' control to determine background luciferase expression

in the luciferase reporter assay (Figure 4.3A). The pGL4 mIL-17 plasmid was considered a suitable positive control, where the mIL-17 promoter was expected to drive luciferase expression (Figure 4.1). The pGL4.73[hRuc/SV40] vector (Promega) encoding the *Renilla reniformis* luciferase reporter gene (*hRluc*) was used as the internal control, which also contains a SV40 early enhancer/promoter to drive expression of *hRluc* (Figure 4.3B). The use of an internal control plasmid was to normalise the values of the experimental plasmids for variation that could be due to transfection efficiency and pipetting errors. The pGL4.10 control (4,242-bp) and pGL4.73[hRuc/SV40] internal control (3,921-bp) plasmids were chosen based on comparable sizes to the constructed pGL4/*IMPG2*-promoter plasmid (5,710-bp) aiming to eliminate plasmid size as a variable in the transfection experiments.

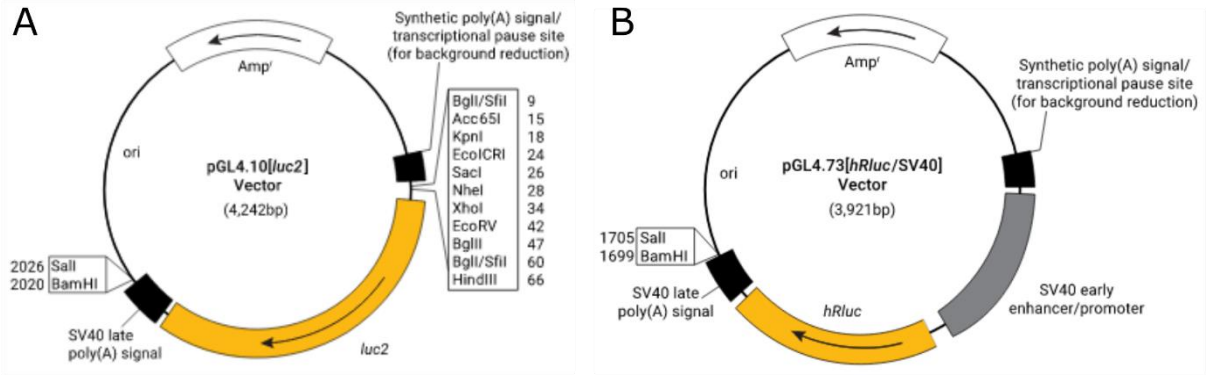


Figure 4.3: Control plasmid maps. **(A)** pGL4.10[*luc2*] basic vector (Promega UK Ltd., Southampton, UK) with no promoter that encodes the luciferase reporter gene *luc2* was used as a background control for luciferase expression. **(B)** pGL4.73[hRuc/SV40] vector (Promega) encoding the *Renilla reniformis* luciferase reporter gene (*hRluc*) and containing a SV40 early enhancer/promoter to drive expression of *hRluc* was used as the internal control.

Transfection and cell lysis

To determine the number of cells to split across 24-well plates, cell numbers were counted manually using a haemocytometer (Neubauer chamber). Two lines of canine skin fibroblast cells cultured in cell culture growth medium at 37 °C were seeded at 5.5 × 10^4 cells per well (canine skin fibroblast cell line A) and 6.5 × 10^4 cells per well (canine skin fibroblast cell line B) at passage three. Human embryonic kidney 293T (HEK239T) cells cultured in cell culture growth medium at 37 °C were seeded at 5 × 10^4 cells per well at passage seven to use as a transfection control. Subsequent to transfection optimisation to determine the optimum ratio of FuGENE®6 transfection reagent (Promega) to DNA, a standard transfection protocol was followed, as summarised in Figure 4.4. All cell lines were plated in single wells (1.9 cm^2) in a 24-well plate using 0.5 mL cell culture growth medium without penicillin and streptomycin in preparation for transfection. Confluent cells were co-transfected the following day by combining 20 µL OptiMEM reduced serum medium (Thermo Fisher Scientific) and 0.9 µL FuGENE®6 transfection reagent (Promega) in a tube and mixed by gently flicking. FuGENE®6 is a lipid-based transfection reagent

and should not be pipetted up and down. Plasmid DNAs were diluted to 200 ng/µL and added to the FuGENE®6/OptiMEM in a 1:20 ratio of experimental plasmid to internal control plasmid and incubated at room temperature for 15 minutes. The FuGENE®6-DNA complex was added to each well of the cell culture plate and mixed by gently swirling. Depending on the confluency of each well, two or three biological replicates were transfected for each canine skin fibroblast cell line to examine after incubation at 37 °C at three time points: 24 hours, 48 hours and 72 hours post-transfection. Following incubation, cells were lysed using Glo lysis buffer (Promega) following manufacturer's instructions. Each lysate was transferred to an Eppendorf tube and stored at -80 °C until ready to conduct the luciferase reporter assay.

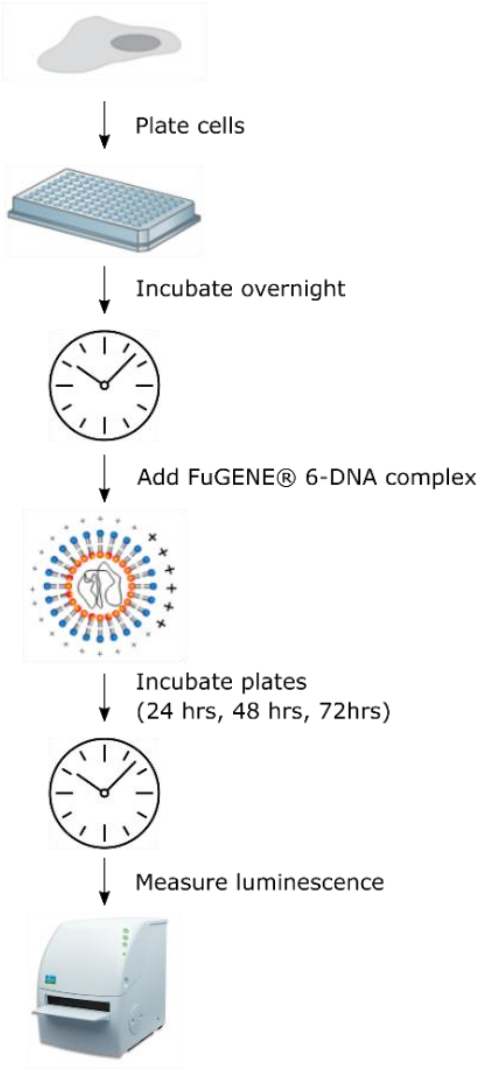


Figure 4.4: Schematic diagram showing steps in a standard transfection protocol.

A green fluorescent protein (GFP) reporter gene control transfection was performed alongside experimental transfections to calculate transfection efficiencies in each cell line. At each time point, media was removed from each well and cells briefly washed in 500 µL PBS (Gibco, Thermo Fisher Scientific). Cells were detached using 250 µL Trypsin with EDTA 0.5% (1X, Invitrogen,

Thermo Fisher Scientific) for 3-5 minutes at room temperature followed by the addition of 250 µL cell culture growth medium without penicillin and streptomycin. To estimate the cell count and estimate transfection efficiency of canine skin fibroblast cell line A, GFP positive cells and a total cell count were estimated manually using a haemocytometer (Neubauer chamber). To determine transfection efficiency of GFP reporter gene transfections by cell quantification using flow cytometry on two biological replicates of canine skin fibroblast cell line B and HEK293T cells, 200 µL of cell lysate was transferred to a 96 well V-bottom assay plate (Corning, Flintshire, UK). In each measurement, to allow for all GFP positive cells to be counted, a maximum of 100,000 events were recorded on an ACEA NovoSampler Pro 2000R NovoCyte Flow Cytometer (ACEA Biosciences Inc., San Diego, USA) equipped with a blue laser (480 nm) and red laser (640 nm) for excitation. GFP reporter gene signal was detected using the fluorescein isothiocyanate (FITC) detection channel (530/30 nm filter with the pass-band centred on 530 nm, and the width of the pass-band is 30 nm) for emission. A medium flow rate was set at a stop condition of 150 µL, and the following mixing parameters were used: speed = 1,200 rpm; duration = 10 seconds; acceleration = 3 seconds. Data were extracted using the ACEA NovoExpress® software (ACEA Biosciences) and analysed as follows. Gating was applied manually to the flow cytometry data using forward scatter (FSC) versus side scatter (SSC) heights to find viable, single cell events and exclude dead cells. Viable events were gated to remove doublets (P1) by forward scatter height (FSC-H) and side scatter height (SSC-H). P1 events were gated by FITC height versus SSC-H. Gated events from the no template control (NTC) were overlaid with experimental GFP events to determine positive GFP cells, where events to the right were considered positive.

Measuring luminescence

Frozen lysates were thawed on ice and each *fluc* and *hRluc* activity was measured using the Dual-Glo(R) Luciferase Assay System (Promega) on an EnVision Multilabel Reader (Perkin Elmer, Buckinghamshire, UK). The results, measured in relative light units (RLU), were normalised as the ratio of *fluc* activity to *hRluc* activity by dividing each *fluc* value by the value for *hRluc* to obtain ratios. Relative fold changes of normalised RLU were calculated to the background control plasmid, pGL4.10, by dividing the experiment mean ratio by the mean ratio of pGL4.10 plasmid. The standard deviation of both canine skin fibroblast cell lines at each time point was calculated using the following formula, where x is the sample mean and n is the sample size (n = 2):

$$\sqrt{\frac{\sum(x-\bar{x})^2}{(n-1)}}$$

The standard deviation of canine skin fibroblast cell lines at each time point was calculated and the standard error estimated as the sample standard deviation divided by the square root of the sample size (i.e. number of observations, n = 2).

4.3 Results

4.3.1 GWAS analysis

A GWAS was conducted using 17 PRA cases and 27 controls. After QC filtering, 108,263 SNPs were included for the analysis of 42 dogs (15 cases and 27 controls; call rate ≥ 97%; MAF ≥ 95%; genotype calls ≥ 90%). The level of genome-wide significance, determined by Bonferroni correction, was $p = 4.6 \times 10^{-7}$ (-log P = 6.34). Analysis of GWAS data revealed a genome-wide significant association on chromosome 33 (CANFA33; $-\log_{10} p_{raw} = 2.2 \times 10^{-16}$) (Figure 4.5A). Visualisation of the SNPs on CANFA33 revealed that 40 SNPs passed the Bonferroni level of genome-wide significance (Figure 4.5B), with a single SNP showing the lowest p-value (BICF2G630247609 at position CANFA33: 8,045,162). The signal on CANFA33 remained significant after correcting for multiple testing using the max(T) permutation procedure, with three SNPs (TIGRP2P385447_RS9043116, BICF2G630247609 and BICF2G630247831) positioned under the GWAS signal with identical $-\log_{10}$ p-values ($p_{genome} = 9 \times 10^{-6}$) (Figure 4.6A; Figure 4.6B). The MDS plot showed a similar distribution of cases and controls (Figure 4.7). After correcting for population stratification and sample relatedness, the signal on CANFA33 remained statistically associated ($p = 1.6 \times 10^{-17}$). Q-Q plots suggested potential population stratification with a moderately increased genomic inflation factor ($\lambda = 1.36$), which decreased to baseline ($\lambda = 1.02$) following corrections of the PLINK results using EMMAX (Figure 4.8).

Visualisation of SNPs either side of the most associated SNP (SNP BICF2G630247609; p-value = 2.2×10^{-16}) in cases sharing the disease-associated haplotype identified a disease-associated interval 1.3 Mb in size that was homozygous in 12 of the 15 cases (Figure 4.9).

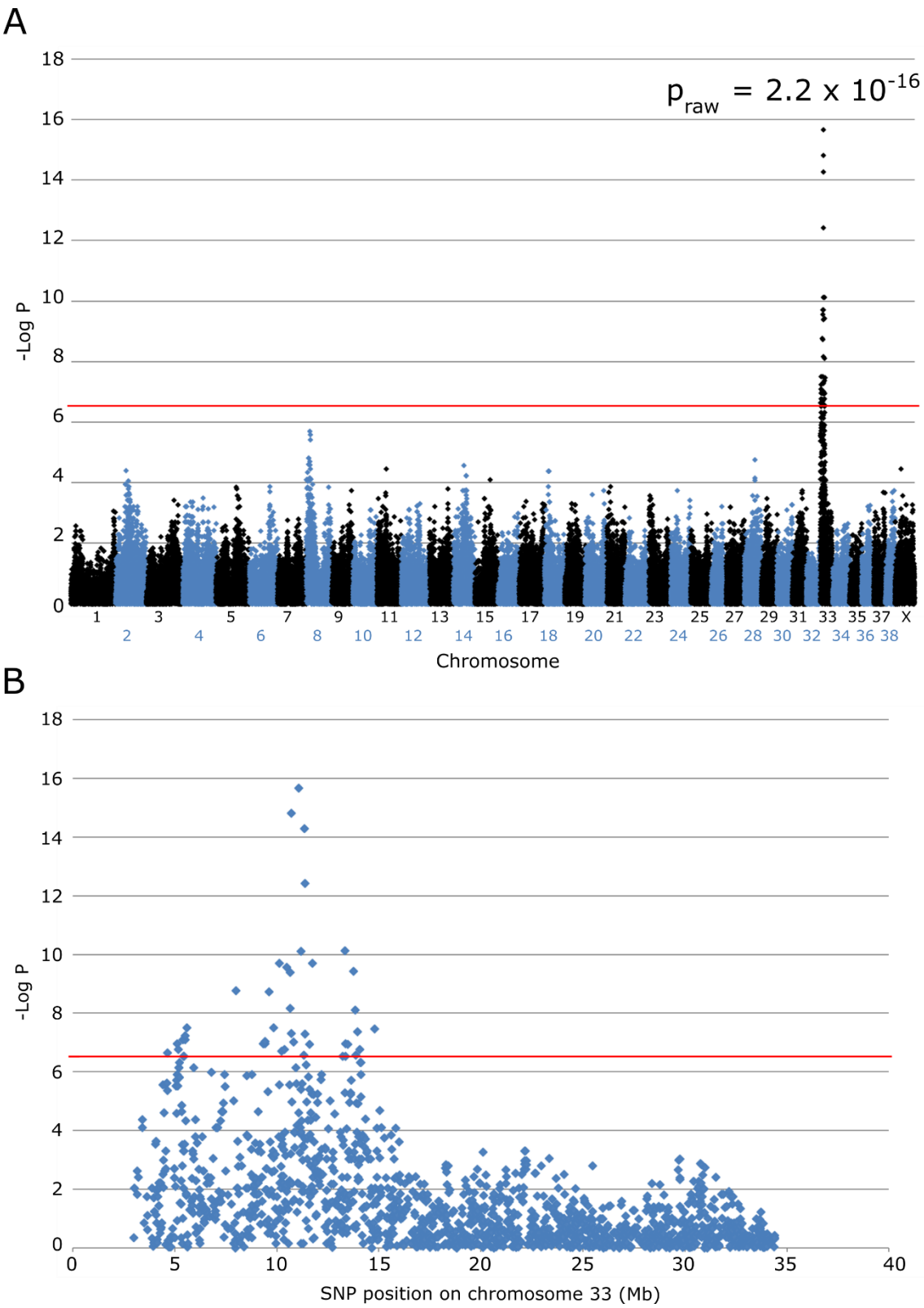


Figure 4.5: Results of the LA PRA GWAS. **(A)** Manhattan plot of unadjusted association analysis showed a statistical association on canine chromosome 33 (CANFA33) (p_{raw} = 2.2×10^{-16}) in LA cases. The level of genome-wide significance, determined by Bonferroni correction, p = 4.6 × 10^{-7} (-log P=6.34), is indicated by a red line. (B)A regional plot of SNPs on CANFA33.

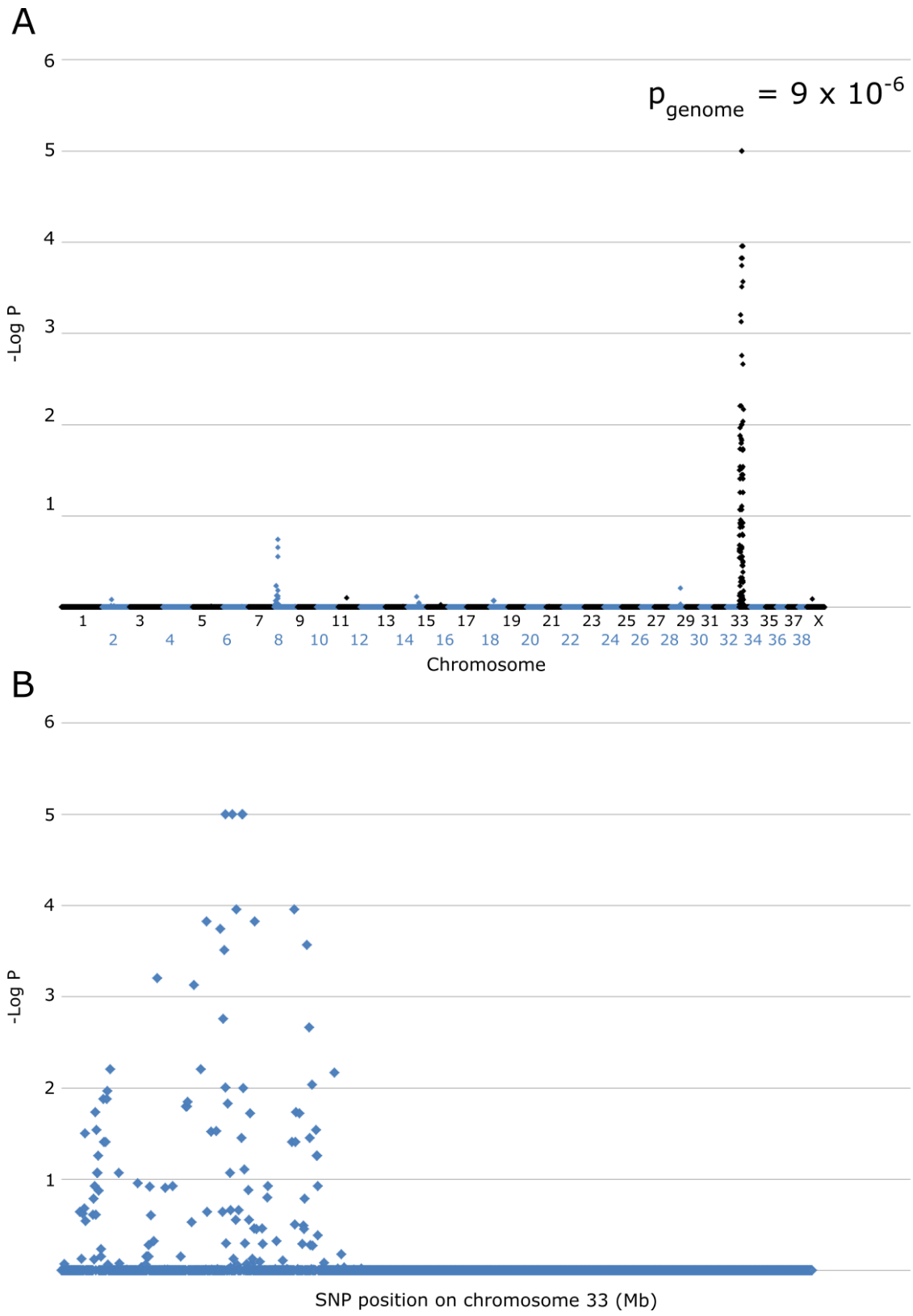


Figure 4.6: Manhattan plots following correcting for multiple testing. **(A)** The signal on canine chromosome 33 (CANFA33) remained significant after correcting for multiple testing using the max(T) permutations procedure (p_{genome} = 9 × 10^{-6}). **(B)** Visualisation of SNP associations on CANFA33 revealed three SNPs with a -$\log_{10}$ p-value of 9 × 10^{-6}.

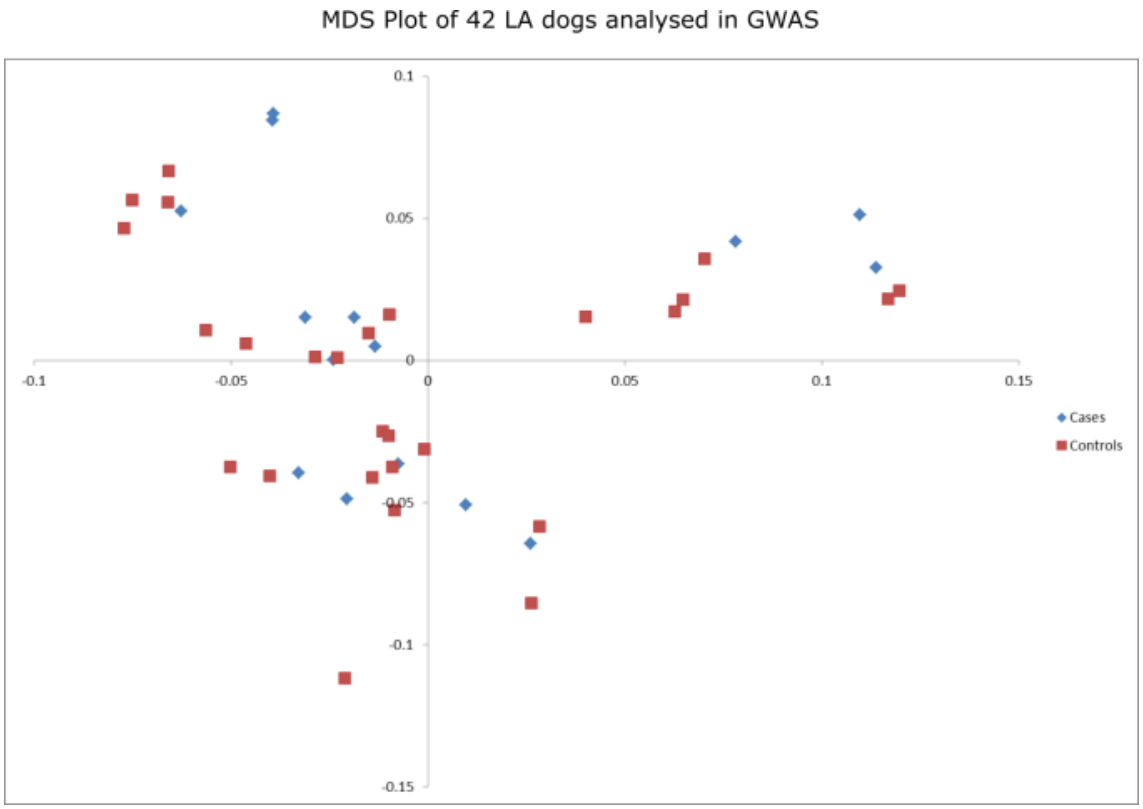


Figure 4.7: A MDS plot to visualise relatedness between the LA case and control sample sets showed a similar distribution of 15 cases and 27 controls analysed in the GWAS.

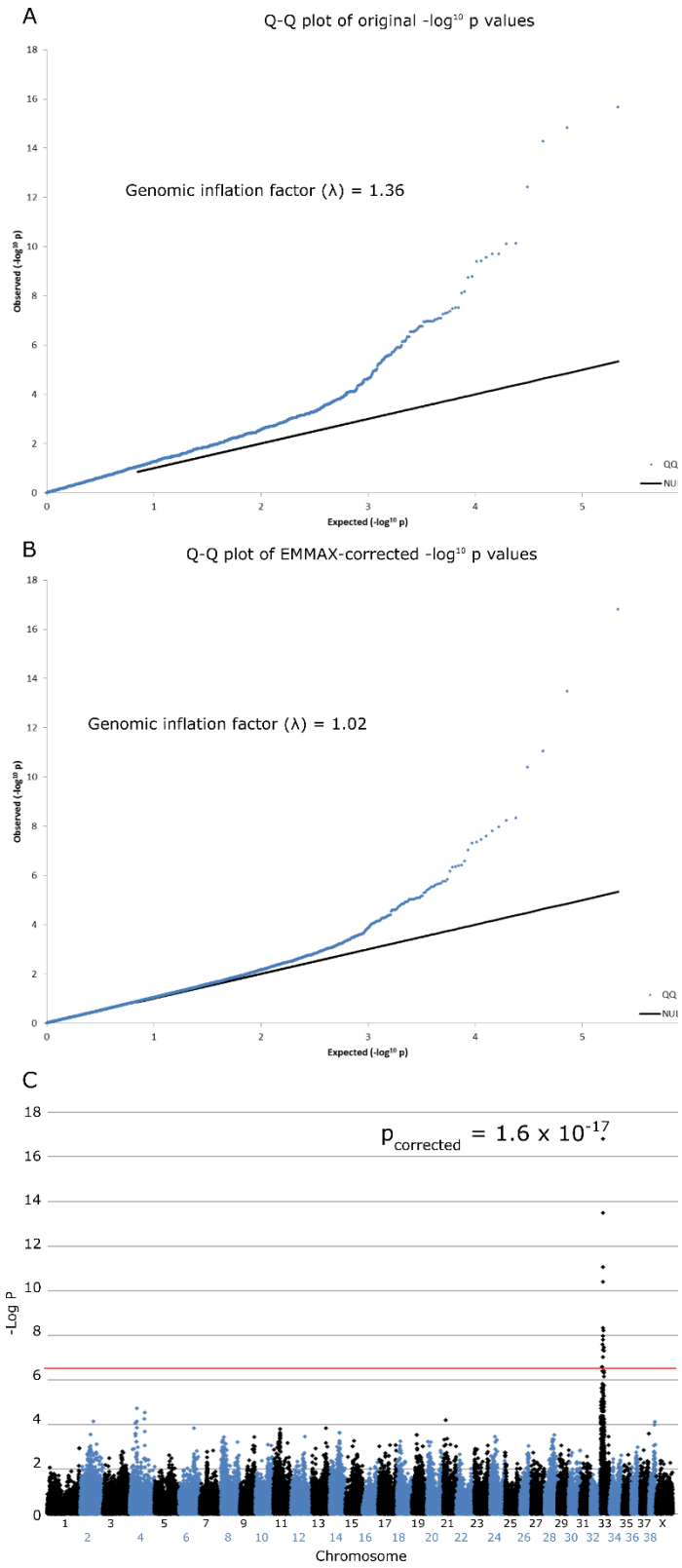


Figure 4.8: Results of the LA PRA GWAS analysis using EMMAX to correct PLINK results . **(A)** The Q-Q plot of the expected and observed –log10 p-values generated from PLINK derived a genomic inflation factor, lambda (λ) = 1.36. **(B)** The Q-Q plot after correcting for population stratification using EMMAX showed a decreased inflation factor, λ = 1.02. **(C)** Manhattan plot of corrected association analysis using EMMAX.

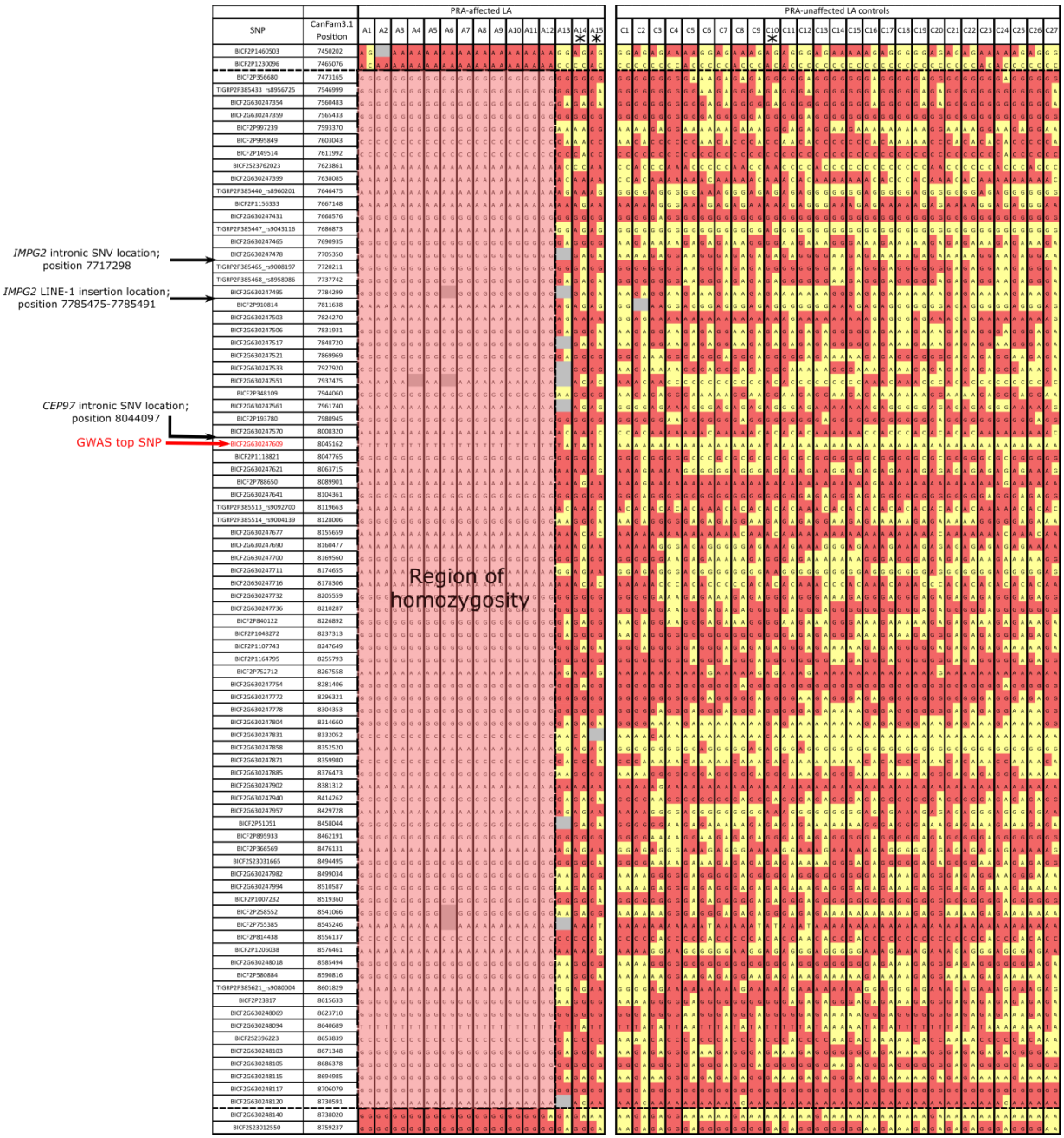


Figure 4.9: Homozygosity analysis of PRA-affected LA dogs after QC filtering. Homozygosity mapping of SNP markers surrounding the most associated SNP from the GWAS (SNP BICF2G630247609; text highlighted in red) in LA PRA cases (A1-15) and controls (C1-27) were used to define the critical region. The yellow coloured boxes represent the reference alleles, the pink coloured boxes highlight the alternate alleles and grey boxes represent missing data. The 1.3 Mb critical region was homozygous in 12 of the 15 cases used in GWAS analysis, as shown by the shaded region between positions CANFA33: 7,465,076-8,738,020. One LA control (C10) is heterozygous for this region, as are two of the three cases that are not homozygous for the critical region, as marked by asterisks (*). Locations of the four variants later followed up after subsequent WGS analysis (LINE-1 insertion, the top associated GWAS SNP (BICF2G630247609), and intronic SNVs in *IMPG2* and *CEP97*) are indicated to show where these variants lie within the critical region.

The defined critical region harbours 21 genes, of which 12 are protein-coding (Table 4.3). Two of these genes were potential candidates: interphotoreceptor matrix proteoglycan 2 (*IMPG2*) and

centrosomal protein 97 (*CEP97*). *IMPG2* has previously been associated with autosomal recessive RP and vitelliform macular dystrophy (VMD) in humans (Bandah-Rozenfeld et al, 2010; Brandl et al, 2017) and is therefore a strong candidate gene for canine PRA. *CEP97* plays a role in centrosome function and cilia formation (Spektor et al, 2007) and although *CEP97* has not directly been implicated with human retinal degenerations, mutations in other centrosomal protein-coding genes have been associated with both syndromic and non-syndromic retinal degenerations (*CEP19*, *CEP78*, *CEP164*, *CEP250* and *CEP290*) (Baala et al, 2007; Chaki et al, 2012; Chang et al, 2006; den Hollander et al, 2006; Frank et al, 2008; Fu et al, 2017; Khateb et al, 2014; Menotti-Raymond et al, 2007; Namburi et al, 2016; Nikopoulos et al, 2016; Valente et al, 2006; Yıldız Bölükbaşı et al, 2018).

Table 4.3: Protein-coding genes situated within the 1.3 Mb critical region. An asterisk (*) highlights genes previously associated with, or within a gene family associated with retinal degeneration in humans.

Gene Name	Abbreviation
ABI family member 3 binding protein	*AB13BP*
Interphotoreceptor matrix proteoglycan 2	*IMPG2**
SUMO1/sentrin specific peptidase 7	*SENP7*
tRNA methyltransferase 10C, mitochondrial RNase P subunit	*TRMT10C*
PEST proteolytic signal containing nuclear protein	*PCNP*
Zinc finger and BTB domain containing 11	*ZBTB11*
Centrosomal protein 97	*CEP97**
Neurexophilin and PC-esterase domain family member 3	*NXPE3*
NFKB inhibitor zeta	*NFKBIZ*
Zona pellucida like domain containing 1	*ZPLD1*
ENSCAFG00000009584	N/A; no human orthologue
Ribosomal protein L24	*RPL24*

A weak signal on chromosome 8 was also present in the GWAS Manhattan plot (Figure 4.5), although this is below the level of genome-wide significance. Visualisation of SNPs spanning the

most associated SNP on chromosome 8 (CANFA8: 10,117,864) showed a small region of homozygosity upstream of the top SNP position (CANFA8: 9,927,621-10,117,864) in four cases, which were also homozygous for the disease-associated region on CANF33, however the homozygous region on CANFA8 was also present in four control dogs. As the cases that were not homozygous for the disease-associated region on CANFA33 could not be accounted for by the chromosome 8 signal, the possibility of this region as another possible PRA locus was excluded.

4.3.2 Secondary GWAS analysis

A secondary GWAS analysis was performed using three LA PRA cases, that were not homozygous for the critical region, and 27 controls. QC filtering resulted in the inclusion of 106,299 SNPs for the analysis of 30 dogs (call rate ≥ 97%; MAF ≥ 95%; genotype calls ≥ 90%). GWAS analysis revealed no genome-wide significant association, as expected due to low statistical power, however, indicative signals on canine chromosome 1 (top SNP BICF2P1007300 at position CANFA1 :66,635,731, chromosome 7 (top SNP BICF2S23052355 at position CANFA7 :70,561,883) and the X chromosome (top SNP BICF2P332508 at position CANFAX :32,952,883) were observed (Figure 4.10). After correcting for multiple testing using the max(T) permutations procedure for 100,000 permutations, these signals remained but were not significant. Further exploration of the genomic regions surrounding these SNPs revealed no suggestive loci associated with disease.

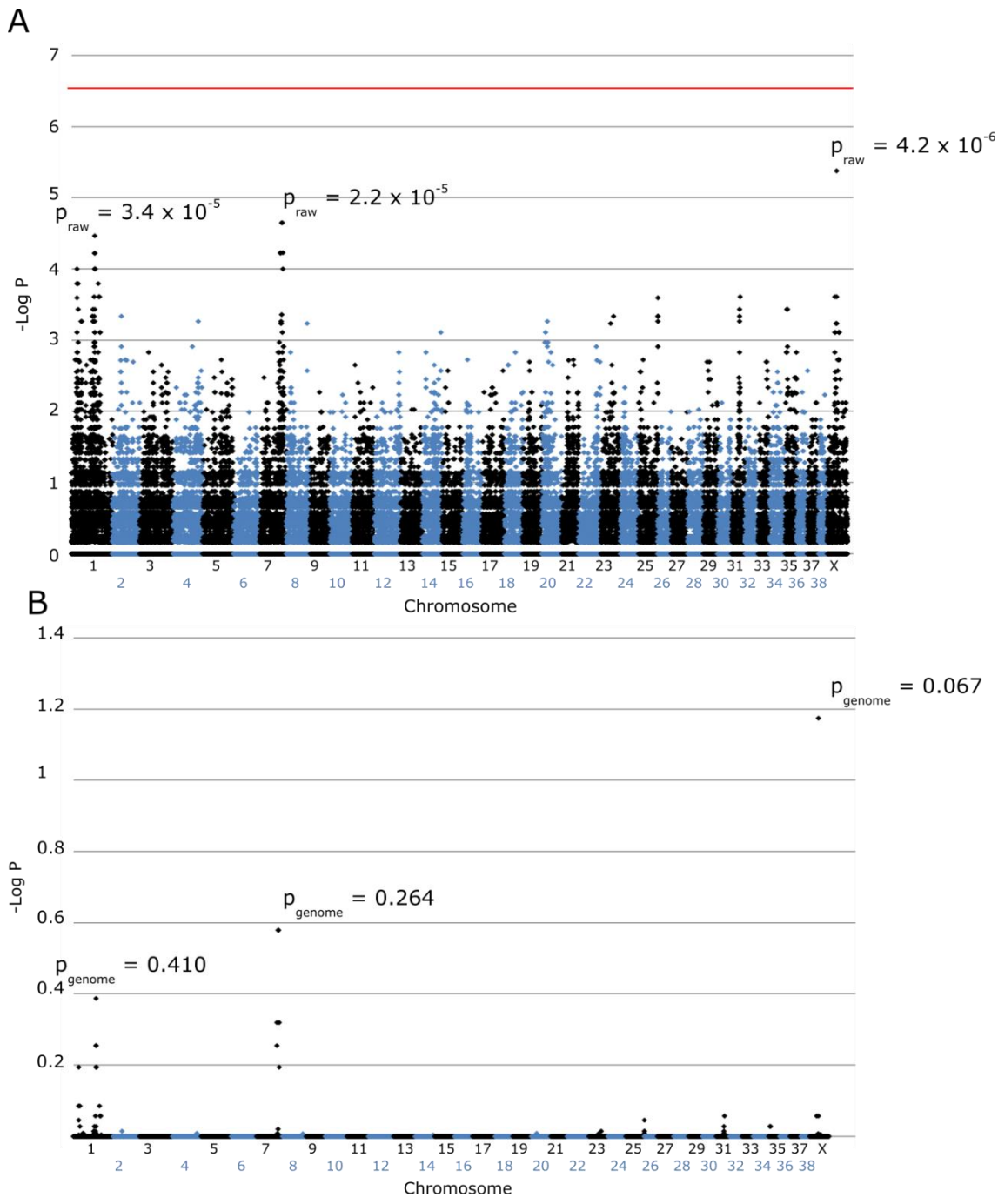


Figure 4.10: Manhattan plots representing -$\log_{10}$ of p-values of LA PRA GWAS using 3 cases and 27 controls showed no statistical associations although weak signals on canine chromosomes 1, 7 and X were observed **(A)**. The level of genome-wide significance, determined by Bonferroni correction, p= 4.6 × 10^{-7} (-log P = 6.34), is indicated by a red line. **(B)** The non-significant signals on chromosomes 1, 7 and X remained after correcting for multiple testing using the max(T) permutations procedure remained weak.

4.3.3 Identification of candidate causal variants underlying the GWAS signal

From examination of WES data previously generated in our laboratory for three LA PRA cases and three LA controls, no exonic or splice site variants that segregated with PRA could be identified. Of these LA dogs used for exome analysis, DNA from one control and all three PRA cases were used for the GWAS in this study. The LA PRA case selected for WGS was also one of the cases in the exome study. The 1.3 Mb homozygous interval was therefore manually interrogated in WGS data of the PRA case using IGV software. A LINE-1 insertion was identified within the critical region in this PRA case, inserted 181-bp upstream of the *IMPG2* gene at CANFA33:7,785,491-7,785,492 (Figure 4.11, track A). The 17-bp target repeat sequence flanking the LINE-1 spans the following coordinates: CANFA33: 7,785,475 - 7,785,491. This insertion was not visible in the WES data of the same PRA case used for WGS analysis (Figure 4.11, track B). In control genomes, the insertion was not present.

Variant filtering of WGS data was performed to highlight additional SNPs and INDELs to consider as candidate variants, intending to highlight those that may be stronger candidates than the LINE-1 insertion. This filtering identified two intronic SNVs situated in retinal candidate genes within the critical region: one in *IMPG2* (G/T SNV; CANFA33:7,717,298) and one in *CEP97* (A/G SNV; CANFA33:8,044,097*)*. The locations of these two intronic SNVs within the defined homozygous critical region are highlighted in Figure 4.9. As this WGS variant filtering prioritised those variants predicted to be highly penetrant, limitations include no further prioritisation of non-coding variants in known IRD/PRA genes, SVs and non-coding variants in potential new candidate genes. The LINE-1 insertion was absent in WGS data from 102 individuals of 52 other breeds and 2 crossbreeds; WGS data from a Hungarian Vizsla dog is shown to illustrate (Figure 4.11, track C). The sequence at each position of the intronic variants were visualised in WGS data from the same 102 canine genomes to determine genotypes. The *IMPG2* intronic SNV was absent in all 102 individuals and the *CEP97* intronic SNV absent in 101 individuals, with one Welsh Springer Spaniel dog identified as heterozygous for the SNV.

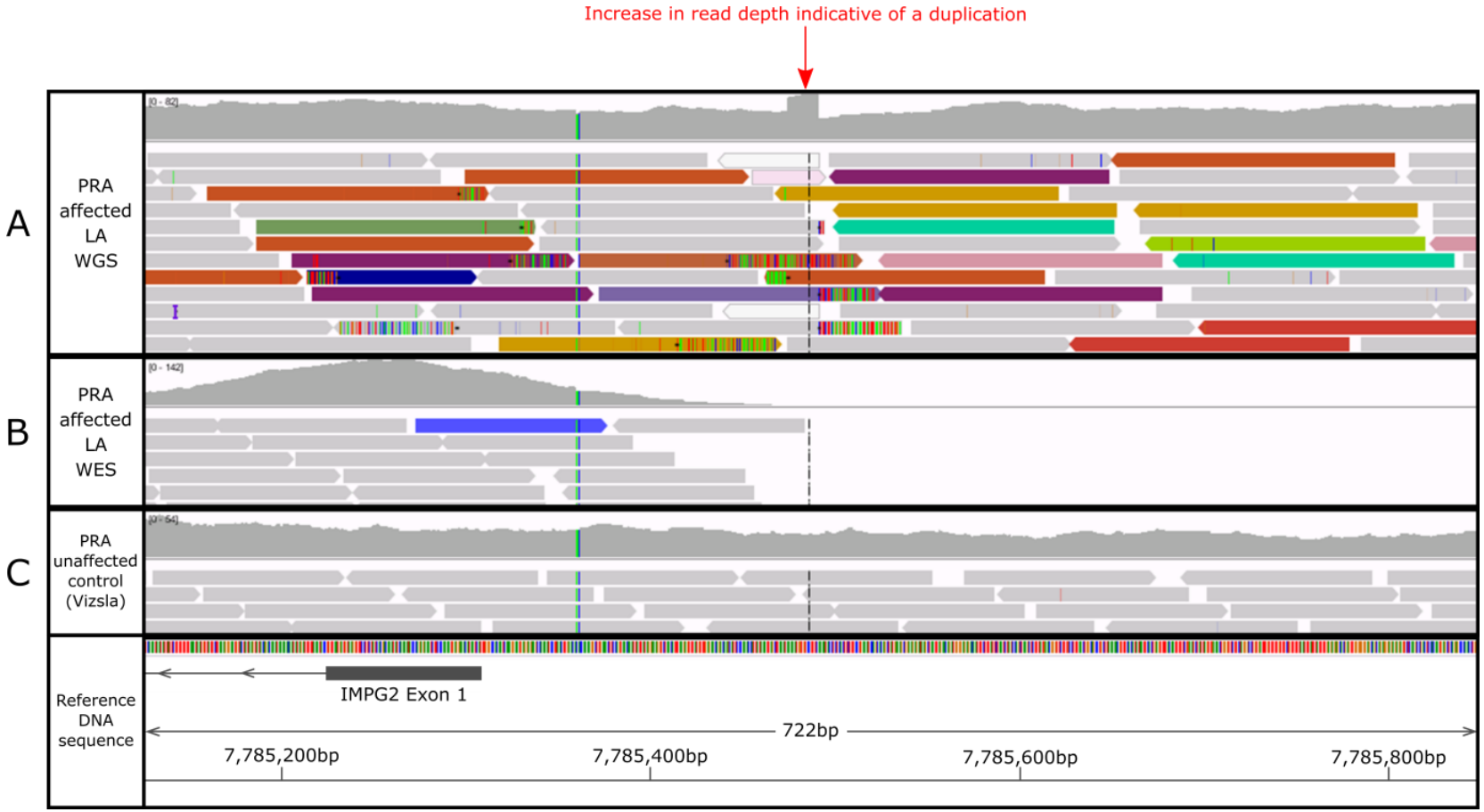


Figure 4.11: IGV display of the LINE-1 insertion upstream of *IMPG2*. WGS (track **A**) and WES (track **B**) alignments over a 722-bp region on CANFA33 from a LA PRA case compared to control WGS data from a PRA non-affected Hungarian Vizsla dog (example, track **C**). The identified LINE-1 insertion is upstream of the retinal candidate gene *IMPG2*. Grey sequencing reads show those aligning normally to the CanFam3.1 canine reference genome at this position. Coloured reads indicate that one of the paired sequencing reads aligns to this region on CANFA33, and the mate in this pair of sequencing reads aligns to another chromosome (each colour represents alignment to a different chromosome), consistent with the presence of a repetitive element insertion.

4.3.4 Sequencing the LINE-1 insertion

Amplification by PCR across the LINE-1 insertion in three LA PRA cases and three LA controls suggested a size of 1.5-2 Kb (Figure 4.12). NGS of the LINE-1 region confirmed an insertion of at least 1,600-bp (Appendix 8). The exact length of the poly-A tail could not be determined due to the low complexity of the short sequencing reads generated from the Illumina sequencing. Analysis in L1Base2 (Penzkofer et al, 2017) of the LINE-1 insertion sequence obtained from *de novo* assembly of NGS reads confirmed a partial transposable element where the sequence partially aligns to open reading frame (ORF) 2 of the L1_cf sequence. As the *IMPG2* gene is located on the reverse DNA strand, WGS alignment shows the LINE-1 sequence has been inserted in the reverse orientation to the *IMPG2* gene, comprising the 17-bp target site duplication AAAAATGGTGTCCTTC (Figure 4.13), a truncated 5' end with a partial 1,629-bp ORF2 and a poly-A tail.

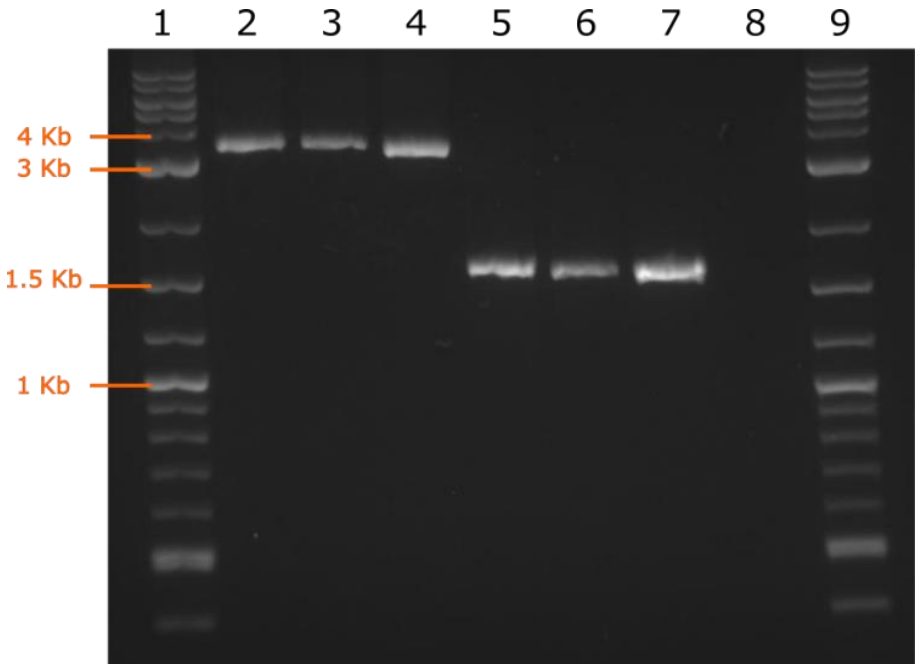


Figure 4.12: Agarose gel electrophoresis of DNA from three LA PRA cases (lanes 2-4), three LA controls (lanes 5-7), a negative control (lane 8) and a 2-log ladder (lanes 1 and 9) (NEB). The expected PCR product size in control LA dogs is 1,606-bp. The product containing the LINE-1 insertion is ~3.5 Kb on the gel, suggesting the LINE-1 insertion is ~1.9 Kb in size.

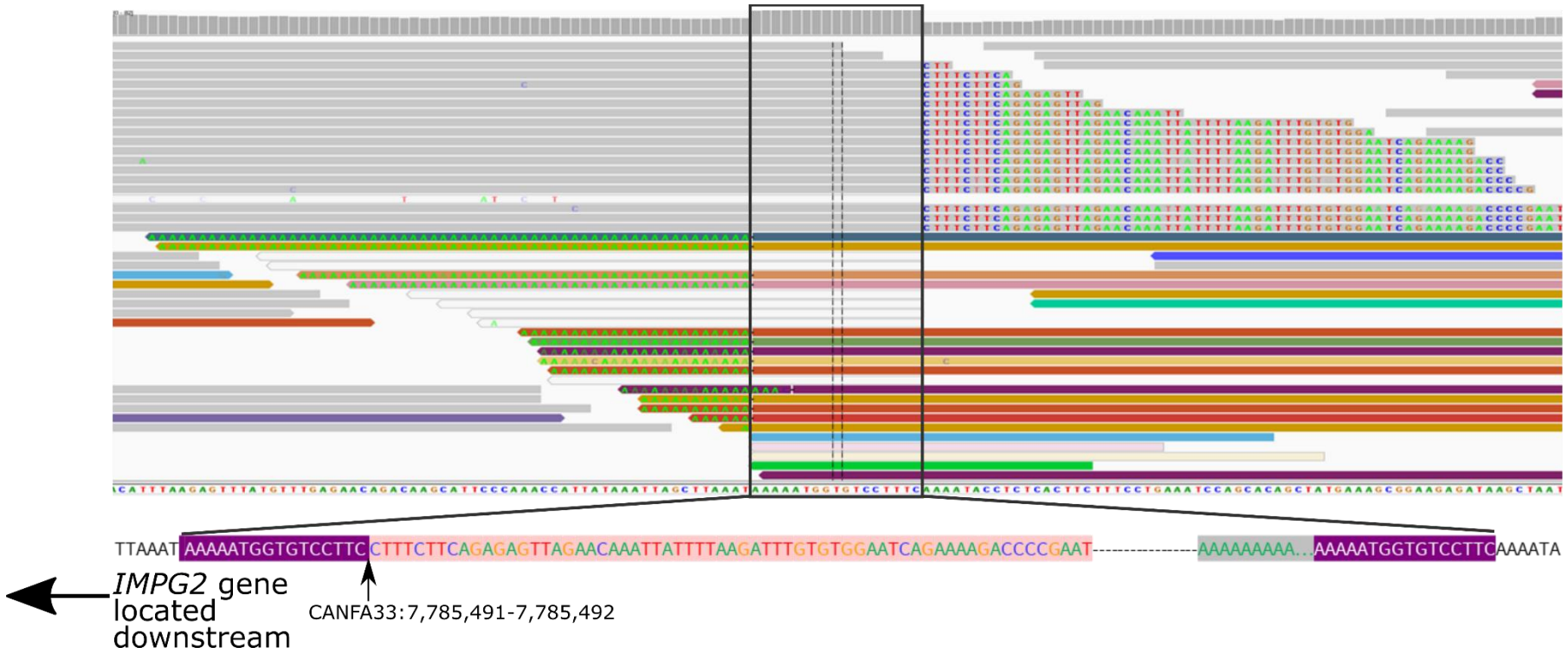


Figure 4.13: Illustration of Illumina sequencing reads in IGV of the LINE-1 insertion in the *IMPG2* gene present in a LA PRA case. The 17-base pair (bp) duplication is highlighted in purple, soft-clipped bases in sequence reads at the start and end of the LINE-1 insertion are boxed in pink and the poly-A tail boxed in grey. The dashed line joining these two sequences represents the remaining 1,472-bp of LINE-1 insertion sequence. The LINE-1 insertion is inserted in the reverse orientation to the *IMPG2* gene, which is located 181-bp downstream of the insertion. Sequence annotation of the LINE-1 insertion is in the forward orientation.

4.3.5 Variant screening

To assess the concordance of the LINE-1 insertion with PRA, an AFLP assay was used to genotype 447 dogs of 123 breeds (Appendix 9) in addition to the cases and controls in the GWAS. Of the individuals in the GWAS dataset analysed, all 12 LA cases that were homozygous for the defined critical region were homozygous for the LINE-1 insertion. In the control set, one heterozygote was present, and the LINE-1 insertion was absent in the other 26 control dogs. In addition, PRA cases of breeds related to the LA (five Shih Tzu dogs, seven Tibetan Spaniels and two Tibetan Terriers), were homozygous for the wild type allele.

Four out of the seventeen PRA-affected individuals included in the original GWAS dataset pre-QC filtering were not homozygous for the 1.3 Mb defined critical region. A secondary GWAS analysis did not yield any suggestive associations. Five additional PRA cases that were not included in the GWAS dataset were available to genotype for the LINE-1 insertion, the top associated SNP from the GWAS (BICF2G630247609) and both intronic SNVs in *IMPG2* and *CEP97*. NNSPLICEv.0.9 (Reese et al, 1997; Berkeley Drosophila Genome Project, 2020) concluded that neither the *IMPG2* or *CEP97* intronic SNVs are located within predicted donor or acceptor splice sites or nearby any splice site predictions, nor do they introduce novel splice sites. In total, 63 LA dogs comprising 22 PRA cases and 41 controls were genotyped for these four variants in an attempt to assess which variant showed the strongest segregation with PRA (Table 4.4). Figure 4.14 shows a schematic diagram of these four variant genotypes across the 63 LA dogs. Five PRA cases (four individuals A13-A16 from the GWAS and individual A22) were either homozygous for the wild type allele or heterozygous across all four variants. Presuming a single-gene disorder model, these five individuals were surmised to be suffering from a genetically different PRA.

Table 4.4: Genotype distributions of the four variants of interest in LA PRA cases and controls. 'Ref' refers to the reference/wild type allele; 'Alt' refers to the alternate allele.

Variant	LA cases			LA controls		
	Alt/Alt	Alt/Ref	Ref/Ref	Alt/Alt	Alt/Ref	Ref/Ref
IMPG2 intronic SNV (chr33:7,717,298)	17	2	3	0	6	35
IMPG2 LINE-1 insertion	17	2	3	0	6	35
CEP97 intronic SNV (chr33:8,044,097)	17	3	2	0	8	33
BICF2G630247609 top GWAS SNP	17	3	2	0	8	33

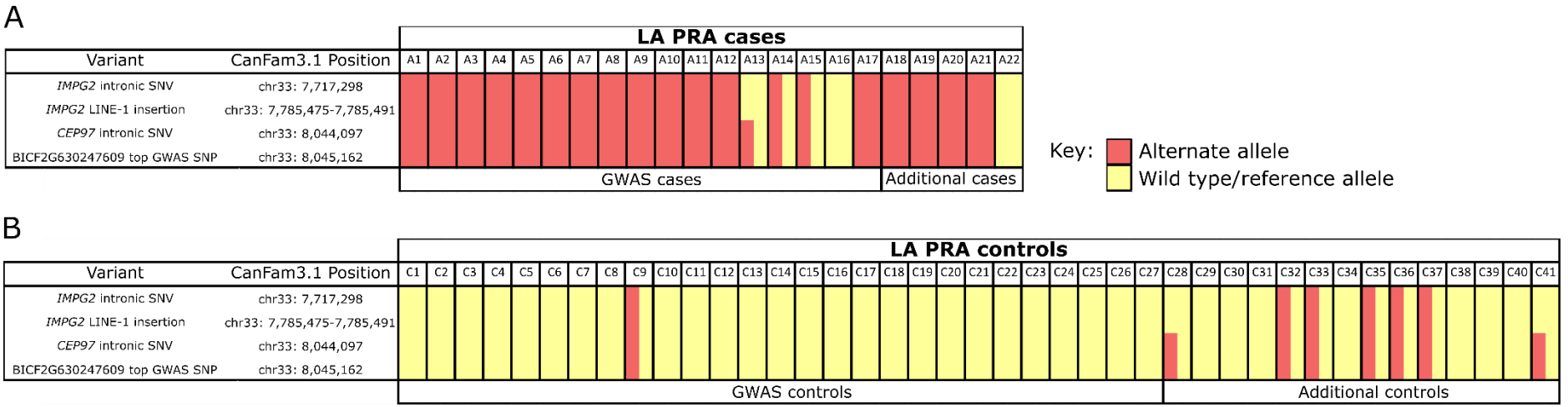


Figure 4.14: Variant genotypes in LA PRA cases and controls. **(A)** A schematic diagram showing genotypes for four variants across 22 LA PRA cases (A1-22) and **(B)** 41 LA controls (C1-41) LA: homozygous alternate allele (coloured pink), homozygous wild type/reference allele (coloured yellow) or heterozygous (coloured pink and yellow).

4.3.6 Promoter and transcription factor binding site (TFBS) predictions

To investigate whether the LINE-1 insertion may disrupt regulation of the *IMPG2* gene, *in silico* analyses of the region surrounding the insertion were performed to search for putative regulatory sequences and promoter sequences. The Gene2Promoter tool suggested that a promoter region exists within 1.5 Kb of the upstream DNA sequence of *IMPG2.* However, when using the PromoterInspector tool to predict eukaryotic Pol II promoter regions in mammalian genome sequences, no such promoter regions were predicted. Analysis of 1.5 Mb upstream and downstream of the LINE-1 insertion breakpoints using the MatInspector tool identified 1,275 matches to putative TFBSs of which 162 were within 150-bp upstream and 150-bp downstream of the LINE-1 breakpoints. Forty-three of these were associated with eye tissue (Appendix 10) including three photoreceptor conserved element 1 TFBSs, one cone-rod homeobox-containing TFBS and one pituitary homeobox 1 TFBS. These five photoreceptor-specific TFBSs belong to the "vertebrates bicoid-like homeodomain transcription factor matrix" family in MatInspector (matrix symbol = V$BCDF) and are located within very close proximity to the LINE-1 insertion (Figure 4.15). All bicoid-like homeodomain TFBSs within this matrix family 'V$BCDF' in the dog are listed in Table 4.5.

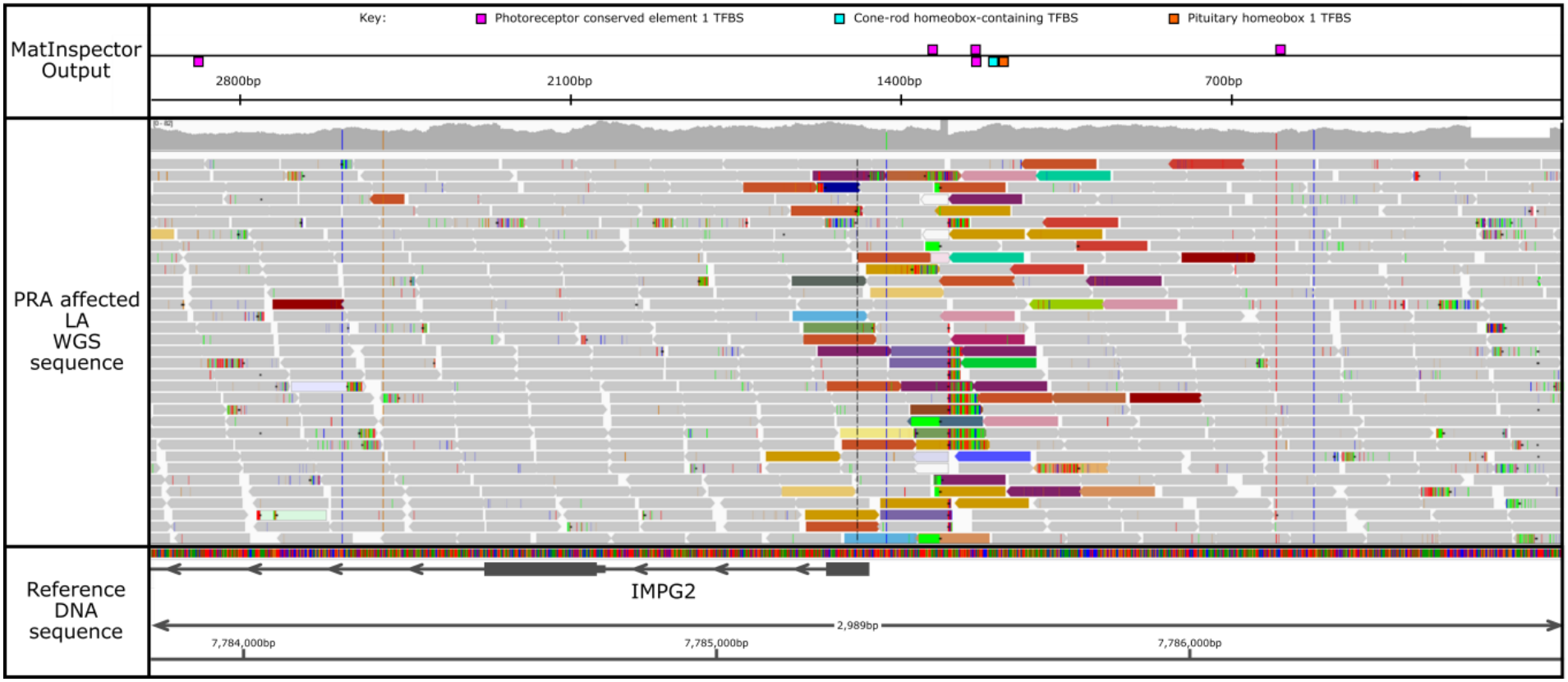


Figure 4.15: Predicted TFBSs surrounding the LINE-1 insertion position. TFBSs within the 'V$BCDF' matrix family from MatInspector within a 2,989-bp region over the LINE-1 insertion are shown in a LA PRA case. Within this matrix family are five photoreceptor conserved element 1 TFBS, one cone-rod homeobox-containing transcription factor/otx-like homeobox TFBS and one pituitary homeobox 1 TFBS.

Table 4.5: Genes encoding transcription factors binding to the TFBSs surrounding the LINE-1 insertion. TFBSs of the 'V$BCDF' matrix family closely situated to the *IMPG2* LINE-1 insertion were identified using MatInspector.

Organism	Genes encoding Transcription Factors
Dog	*CRX*, *DMBX1*, *GSC2*, *OTX1*, *OTX2*, *PITX1*, *PITX2*, *PITX3*
Human	*CRX*, *DMBX1*, *GSC*, *GSC2*, *OTX1*, *OTX2*, *PITX1*, *PITX2*, *PITX3*

4.3.7 Luciferase reporter assay

Given that the LINE-1 insertion is upstream of the *IMPG2* gene, within the predicted promoter, it was hypothesised that the insertion may cause PRA through dysregulation of *IMPG2* expression. To assess this, a luciferase reporter assay was attempted, initially optimising the assay using the non-interrupted *IMPG2* predicted promoter sequence as a positive control to establish that the predicted promoter drives luciferase expression in the generated plasmid when transfected into a canine cell line.

Cell line selection

To determine which cell type to use for the reporter assay, qPCR was conducted using cDNA synthesised from canine skin fibroblast cells, MDCK cells and MSCs, and *IMPG2* expression in these cell types was compared to retinal cDNA synthesised from a control Golden Retriever dog with no ocular phenotypes. The *IMPG2* target expression level was normalised against a housekeeping gene, *TBP,* which is ubiquitously expressed across tissues. R^2 and amplification efficiency values determined from standard curves of each target in control canine retinal cDNA are listed in Table 4.6. qPCR analysis indicated similar levels of *IMPG2* expression in canine skin fibroblast cells, MDCK and MSCs (Figure 4.16), albeit at a much lower expression to that of *IMPG2* in retina (Table 4.7; Figure 4.17). As no retinal cell lines were available for use at the time of study, canine skin fibroblast cells were the cell of choice to perform luciferase reporter assays and assess if the LINE-1 insertion affects *IMPG2* promotor activity.

Table 4.6: Standard curves of qPCR targets ascertained using control canine retinal cDNA.

Target	Slope	R^2	Efficiency (%)
TBP	-3.11	0.994	98.379
IMPG2	-3.41	0.996	96.471

Table 4.7: Expression fold changes of *IMPG2* in various canine cell types. Expression fold change calculated using the delta-delta Ct method (2^-ΔΔCt) of *IMPG2* target versus *TBP* target in canine skin fibroblast cells, canine kidney (MDCK) cells and canine mesenchymal stem cells (MSC).

Cell/Tissue	Expression Fold Change (2^-ΔΔCt)
MSC	0.00055
MDCK	0.00060
Canine Skin Fibroblast cells	0.00033
Retina (Control)	1

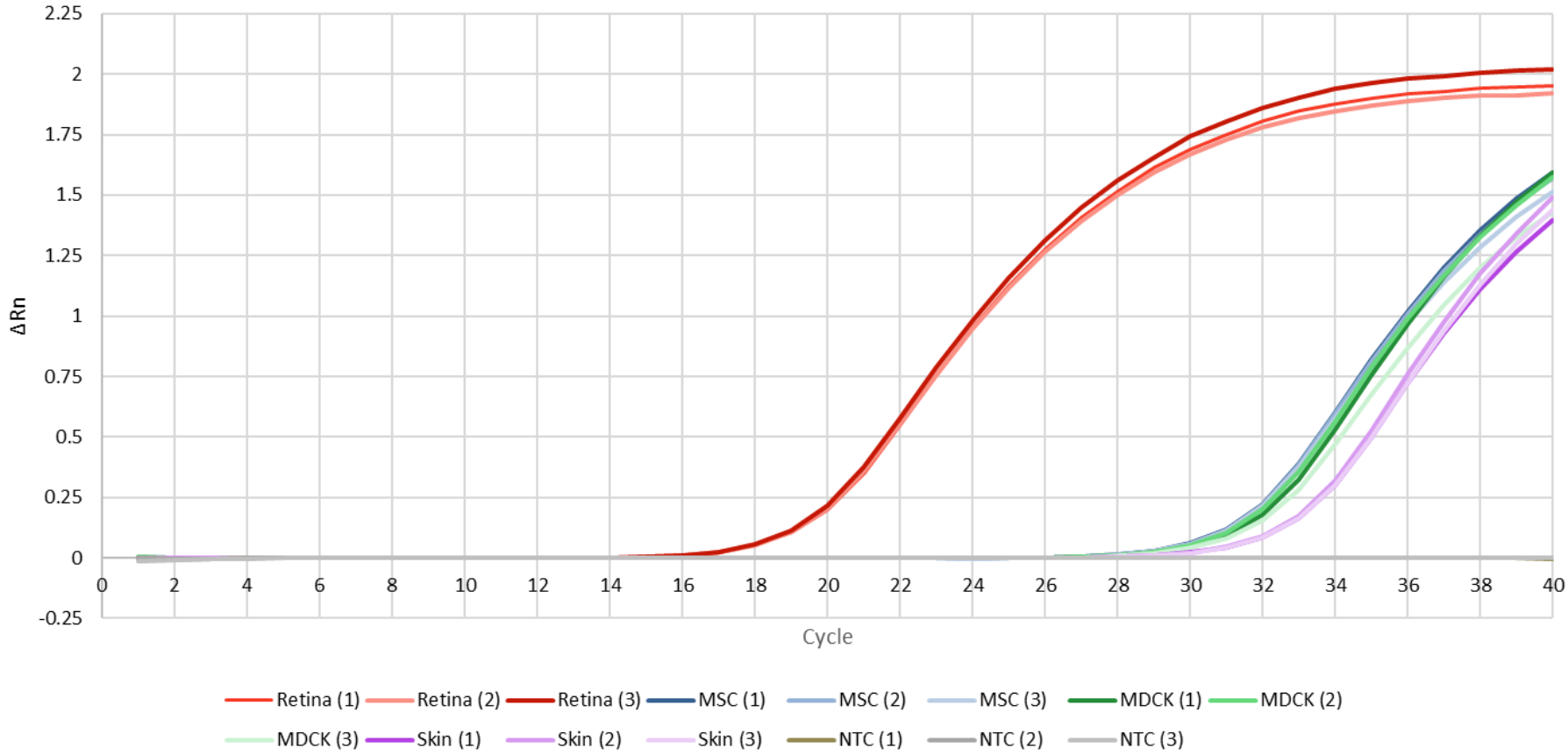


Figure 4.16: *IMPG2* expression in various cell types. A graph to show ΔRn (change in fluorescence) against cycle number for expression of *IMPG2*. cDNA synthesised from RNA extracted from canine retinal tissue was used as a positive control, with *IMPG2* known to be highly expressed in the retina. cDNA synthesised from RNA extracted from MSC, MDCK and canine skin fibroblast cells were assessed for *IMPG2* expression in comparison to the canine cDNA from retinal tissue. MSC, MDCK and canine skin fibroblast cells expressed *IMPG2*, but with very low efficiency, thus canine skin fibroblast cells were chosen to utilise for further reporter assays.

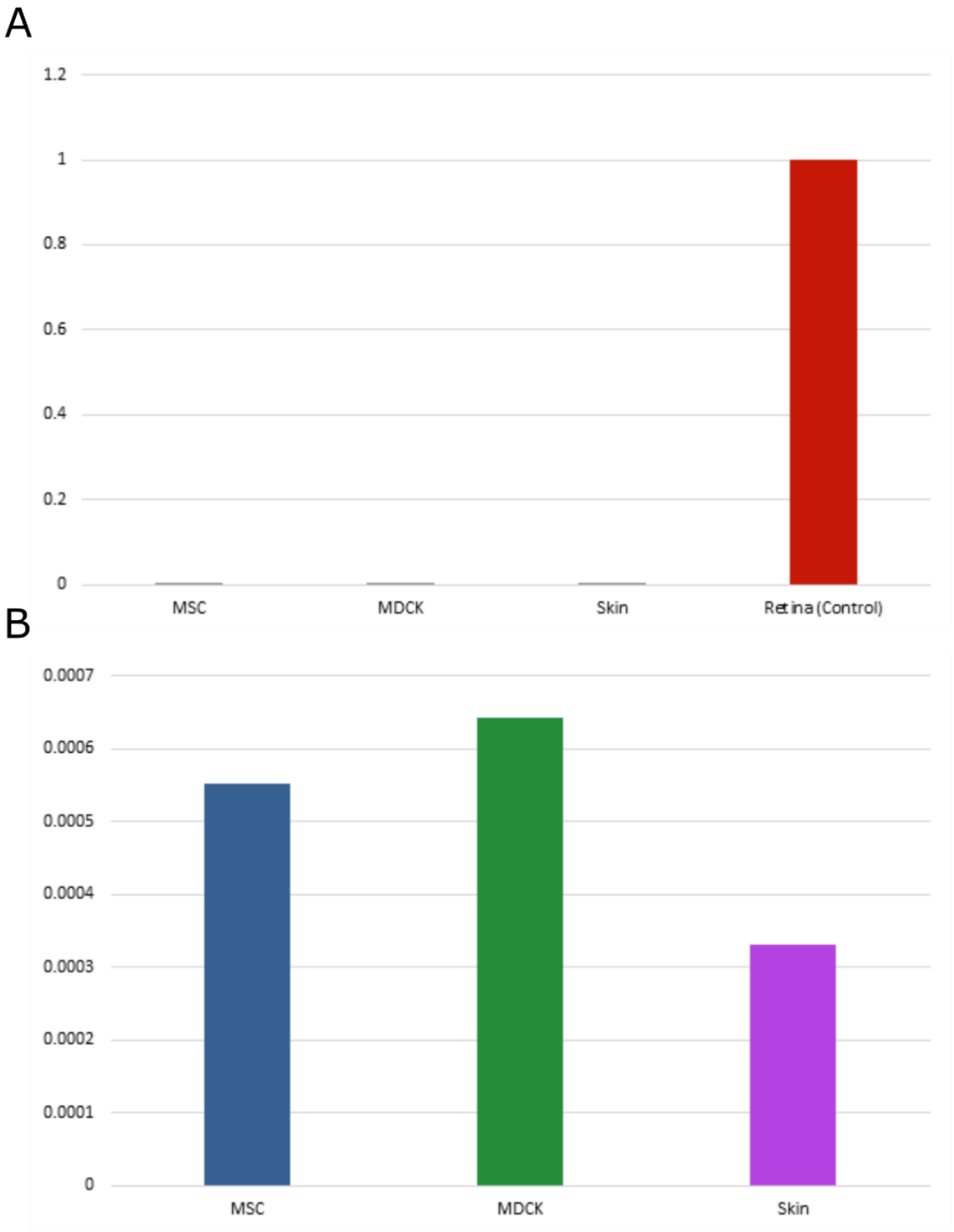


Figure 4.17: Expression fold changes of *IMPG2* in various cell types. **(A)** Expression fold change (2^-ΔΔCt) of *IMPG2* target versus TBP target in canine skin fibroblast cells, canine kidney (MDCK) cells, and canine mesenchymal stem cells (MSC). The retinal cDNA extracted from retinal tissue of a control Golden Retriever dog was used as the baseline value (n = 1). **(B)** Zoom in of canine MSC, MDCK and skin fibroblast cells for expression fold change (2^-ΔΔCt) values.

Generation of the pGL4/*IMPG2*-promoter plasmid

The mIL-17 promoter was removed from the pGL4 mIL-17 plasmid (Figure 4.18), resulting in the pGL4 backbone for the *IMPG2* promoter insert to be incorporated. Restriction enzymes NheI and BglII (both NEB) were used to excise the mIL-17 promoter from the vector. As BglII cut sites were present within the *IMPG2* insert sequence, the BamHI restriction enzyme (NEB), with the same overhang as BglII, was used in restriction digests with the *IMPG2* insert to generate sticky ends that were compatible with the vector enzyme cut sites for ligation. The constructed plasmid was termed 'pGL4/*IMPG2*-promoter'.

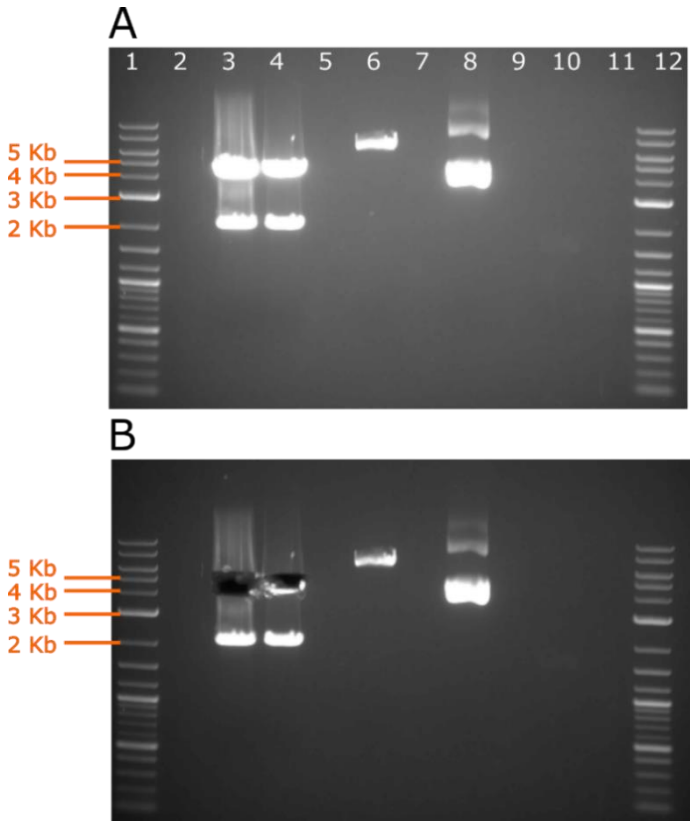


Figure 4.18: Agarose gel electrophoresis following plasmid restriction digestion. **(A)** Pre-excised agarose gel of 50 µL digested pGL4 mIL-17 plasmid was loaded in two wells (A3,A4) on an agarose gel to determine digestion had resulted in the separation of the pGL4 backbone (4,231-bp) and the mouse mIL-17 promoter (2,000-bp). A 2-log ladder (NEB) was run to measure DNA fragment sizes (A1, A12). A single digest of the pGL4 mIL-17 plasmid was loaded to show BglII had successfully digested the plasmid cutting at one position for a single product. An aliquot of uncut pGL4 mIL-17 plasmid was loaded as a control well to show the complete vector in supercoiled and open-circular confirmations, as shown by multiple bands (A8). **(B)** Agarose gel post-excision of the 4,231-bp pGL4 backbone.

To determine successful ligation of the vector and insert, and to verify positive transformation and purification of the pGL4/*IMPG2*-promoter plasmid, a restriction digest and Sanger sequencing were performed on two positive colonies from the transformation. A restriction digest using NheI and PspOMI on two preparations of the plasmid indicated DNA fragments were within expected size ranges (two products expected at 4,143-bp and 1,567-bp) (Figure 4.19).

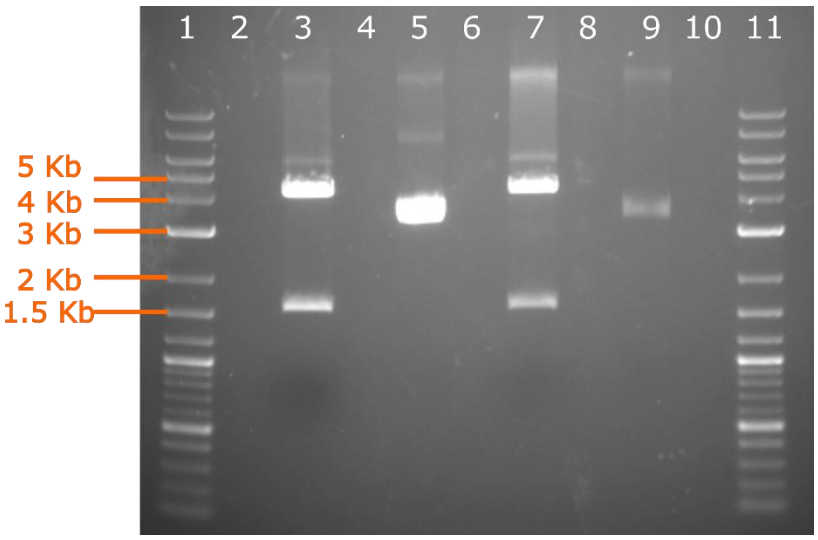


Figure 4.19: Agarose gel electrophoresis to verify plasmids. Two positive colonies from transformations that underwent plasmid purification were digested using NheI and PspOMI enzymes to ensure correct generation of the pGL4/*IMPG2*-promoter plasmid. Digested plasmids were assessed on a 0.8% agarose gel (wells 3 and 7) against undigested plasmid (wells 5 and 9) against a 2-log ladder (wells 1 and 11) indicating DNA fragments within the expected size ranges (4,143-bp and 1,567-bp).

Sanger sequencing (using primers to amplify the entire 5,710-bp plasmid) indicated that the *IMPG2*-promoter sequence had successfully inserted into the pGL4 backbone, generating the pGL4/*IMPG2*-promoter plasmid in the two colonies. As restriction enzymes NheI and BglII/BamHI (backbone vector/insert) were used in digestions of the backbone and insert, these restriction enzyme site sequences were used to determine positions where the insert sequence and pGL4 backbone sequence were ligated (Figure 4.20).

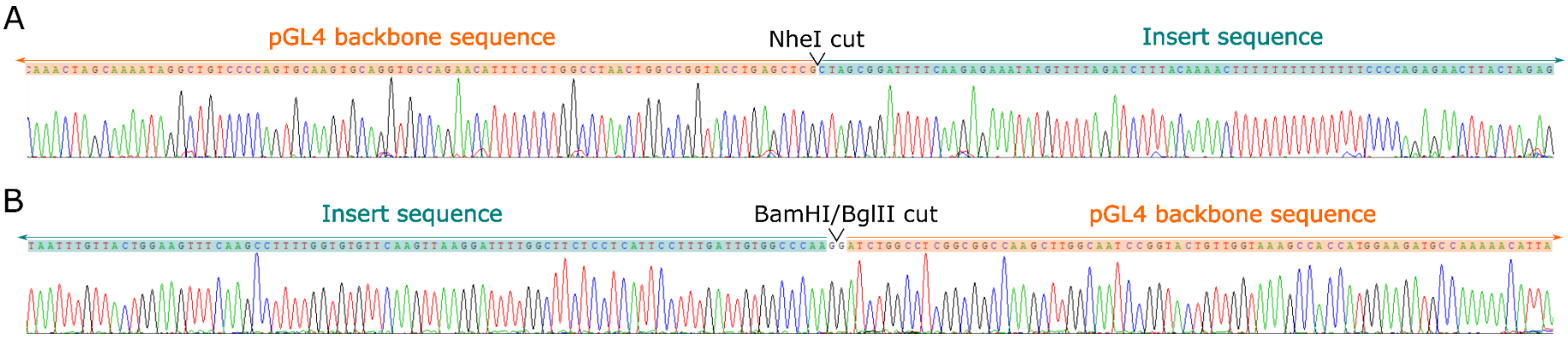


Figure 4.20: Sanger sequencing chromatograms of the pGL4/*IMPG2*-promoter plasmid DNA containing the cloned *IMPG2*-promoter insert in the pGL4 backbone spanning the positions of ligation at **(A)** the NheI enzyme cut site and **(B)** the BamHI/BglII enzyme cut site. As restrictions enzymes NheI and BglII/BamHI (backbone vector/insert) were used, these restriction enzyme site sequences could be used to determine positions where the insert sequence and pGL4 backbone sequence were ligated.

Transfection efficiency

Alongside experimental transfections, transfections using the GFP reporter gene were performed in order to monitor the transfection efficiency of each cell line. In addition to the use of canine skin fibroblast cell lines, a highly transfectable cell line (HEK293T) was used as a control to determine positive or negative GFP transfection using FuGENE®6 transfection reagent. HEK293T cells have shown high transfection efficiencies and high expression levels in luciferase assays using FuGENE®6 (technical notes, Promega). GFP detection was monitored manually by cell counting for canine skin fibroblast cell line A, and using flow cytometry analyses on canine skin fibroblast cell line B and HEK293T cells, using a NTC for each line for normalisation. Transfection efficiencies were calculated (Table 4.8) and flow cytometry results for GFP transfections in canine skin fibroblast cell line B and HEK293T cells are presented in Figure 4.21. The HEK293T cells showed high GFP mean transfection efficiencies: 95.56% and 89.16% at 24 hours and 48 hours post-transfection, respectively. Mean GFP transfection efficiencies of canine skin fibroblast cells were much lower, ranging from 0.89 – 9.47% at 24 hours post-transfection, with the highest GFP transfection efficiency of 26.04% in the canine fibroblast cell line B measured at 48 hours post-transfection.

Table 4.8: Transfection efficiencies using the GFP reporter gene. Transfection efficiencies (%) of canine skin fibroblast cell line A transfected with GFP plasmid DNA were estimated 24 hours post-transfection by performing a manual cell count of the number of fluorescent positive cells divided by the total cell count estimated using a haemocytometer. GFP transfection efficiencies were calculated using flow cytometry in canine skin fibroblast cell line B 24 hours, 48 hours and 72 hours post-transfection, and 24 hours and 48 hours post-transfection in HEK293T cells. All GFP transfections were performed in duplicate using FuGENE®6 transfection reagent (Promega). Mean transfection efficiencies were calculated for each cell line and the standard deviation reported.

Replicate	Cell count	Total number of cells	Transfection efficiency (%)	Mean transfection efficiency (%)	Standard deviation
Canine skin fibroblast cell line A					
24 hrs post transfection					
1	434	80,000	0.54		
2	990	80,000	1.24	0.89	0.49
Canine skin fibroblast cell line B					
24 hrs post transfection					
1	1,376	15,255	9.02		
2	1,463	14,737	9.93	9.47	0.64
48 hrs post transfection					
1	2,264	9,069	24.96		
2	1,797	6,628	27.11	26.04	1.52
72 hrs post transfection					
1	2,772	11,000	25.20		
2	3,949	17,885	22.08	23.64	2.21
HEK293T cells					
24 hrs post transfection					
1	77,029	79,707	96.64		
2	67,041	70,951	94.49	95.56	1.52
48 hrs post transfection					
1	70,618	78,384	90.09		
2	60,777	68,890	88.22	89.16	1.32

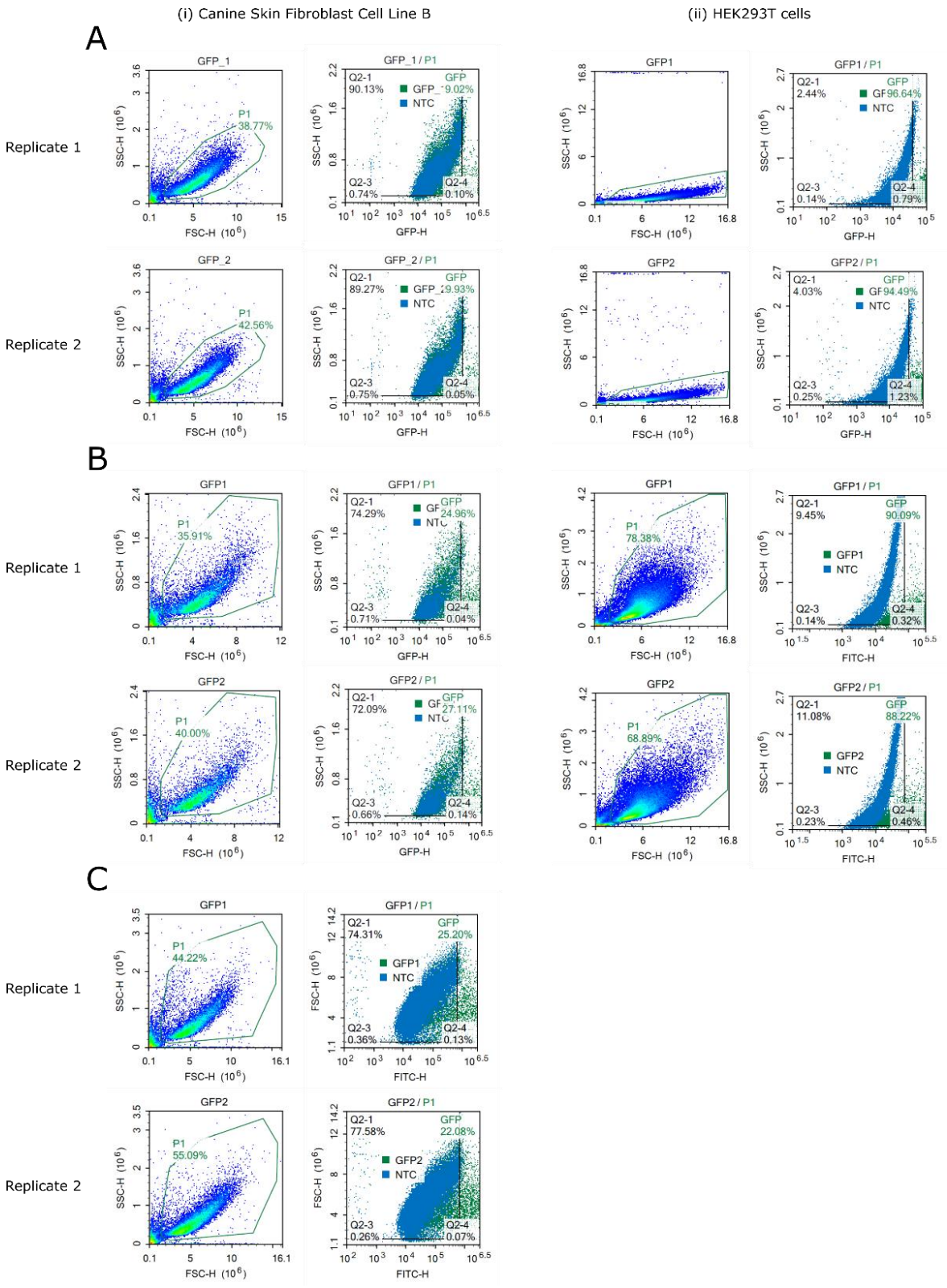


Figure 4.21: Forward and side scatter density plots for GFP transfected cells. Flow cytometry analysis for two biological replicates of (i) canine skin fibroblast cell line B and (ii) HEK293T cells at three time points post transfection using a GFP reporter gene: **(A)** 24 hours, **(B)** 48 hours and **(C)** 72 hours (not applicable for HEK293T cells). P1 indicates gating of the viable events to remove doublets by forward scatter height (FSC-H) and side scatter height (SSC-H). P1 events were gated by FITC height versus SSC-H. Gated events from the control (NTC; blue dots) were overlaid with experimental GFP events (green dots) to determine positive GFP cells, where events to the right were considered positive.

Dual-luciferase reporter assay

Luciferase activity was measured on a luminometer in RLU. RLU values were normalised in individual wells to obtain ratios. As each construct was assayed in duplicate or triplicate wells, a mean normalised *fluc/hRluc* ratio was determined. The relative expression of *fluc* to *hRluc* of the background control (pGL4.10) was used as a baseline fold change (n = 1) and other constructs compared to the activity of pGL4.10 to determine normalised fold changes in activity. Raw RLU results, mean ratios and calculated fold changes for each plasmid and cell line are listed in Table 4.9 and Table 4.10. For the two canine skin fibroblast cell lines, mean *fluc*/*hRluc* ratios were calculated and a fold change generated relative to pGL4.10 values for each cell line and each time point. A mean fold change for each plasmid experiment, at each time point, for the two canine skin fibroblast cell lines and the HEK293T cells was calculated (Table 4.11). The standard deviation of canine skin fibroblast cell lines at each time point was calculated, and the standard error ascertained from the standard deviation with two incidences (cell line A and cell line B).

In the canine skin fibroblast cell lines, the pGL4 mIL-17 plasmid shows a higher fold change than the control pGL4.10 plasmid. In the HEK293T cells, pGL4 mIL-17 is not highly expressed. The pGL4/*IMPG2*-promoter construct shows a lower fold change relative to pGL4.10 in both canine skin fibroblast cell lines (Figure 4.22) and HEK239T cells (Figure 4.23), suggesting that the *IMPG2* promoter is not active in these cell types.

Table 4.9: Raw luminometer readings for two lines of canine skin fibroblast cells (line A, line B) transfected with each plasmid measuring firefly (F) and Renilla (R) luciferase activity in relative light units (RLU).

pGL4.10						pGL4/IMPG2						pGL4 mIL-17					
Time Point	**Test Replicate**	**Luciferase activity (RLU)**		**Ratio (F/R)**	**Relative Fold Change**	**Time Point**	**Test Replicate**	**Luciferase activity (RLU)**		**Ratio (F/R)**	**Relative Fold Change**	**Time Point**	**Test Replicate**	**Luciferase activity (RLU)**		**Ratio (F/R)**	**Relative Fold Change**
		Firefly (F)	**Renilla (R)**					**Firefly (F)**	**Renilla (R)**					**Firefly (F)**	**Renilla (R)**		
24 hrs	Line A 1	4	192	0.021		24 hrs	Line A 1	8	460	0.017		24 hrs	Line A 1	32	200	0.160	
	Line A 2	8	208	0.038			Line A 2	4	500	0.008			Line A 2	28	264	0.106	
	Line A 3	4	216	0.019			Line A 3	8	452	0.018							
		Mean ratio		**0.026**	**1.00**			**Mean ratio**		**0.014**	**0.55**			**Mean ratio**		**0.133**	**5.13**
	Line B 1	4	112	0.036			Line B 1	0	76	0.000			Line B 1	24	172	0.140	
	Line B 2	4	124	0.032			Line B 2	4	72	0.056			Line B 2	24	124	0.194	
		Mean ratio		**0.034**	**1.00**			**Mean ratio**		**0.028**	**0.82**			**Mean ratio**		**0.167**	**4.90**
48 hrs	Line A 1	44	1080	0.041		48 hrs	Line A 1	12	1500	0.008		48 hrs	Line A 1	104	1788	0.058	
	Line A 2	36	1240	0.029			Line A 2	20	1632	0.012			Line A 2	92	1576	0.058	
	Line A 3	32	1056	0.030			Line A 3	20	1684	0.012							
		Mean ratio		**0.033**	**1.00**			**Mean ratio**		**0.011**	**0.32**			**Mean ratio**		**0.058**	**1.75**
	Line B 1	60	1696	0.035			Line B 1	20	1096	0.018			Line B 1	172	1628	0.106	
	Line B 2	80	1568	0.051			Line B 2	32	1160	0.028			Line B 2	176	1336	0.132	
		Mean ratio		**0.043**	**1.00**			**Mean ratio**		0.023	**0.53**			**Mean ratio**		**0.119**	**2.75**

Table 4.9 (continued).

pGL4.10						pGL4/*IMPG2*						pGL4 mIL-17					
Time Point	**Test Replicate**	**Luciferase activity (RLU)**		**Ratio (F/R)**	**Relative Fold Change**	**Time Point**	**Test Replicate**	**Luciferase activity (RLU)**		**Ratio (F/R)**	**Relative Fold Change**	**Time Point**	**Test Replicate**	**Luciferase activity (RLU)**		**Ratio (F/R)**	**Relative Fold Change**
		Firefly (F)	**Renilla (R)**					**Firefly (F)**	**Renilla (R)**					**Firefly (F)**	**Renilla (R)**		
72 hrs	Line A 1	76	896	0.085			Line A 1	24	860	0.028							
	Line A 2	68	748	0.091		72 hrs	Line A 2	28	1208	0.023		72 hrs	Line A 1	232	1844	0.126	
	Line A 3	88	1252	0.070			Line A 3	40	1340	0.030			Line A 2	216	1708	0.126	
		Mean ratio		**0.082**	**1.00**			**Mean ratio**		**0.027**	**0.33**			**Mean ratio**		**0.126**	**1.54**
	Line B 1	108	884	0.122			Line B 1	36	572	0.063			Line B 1	268	1568	0.171	
	Line B 2	108	2076	0.052			Line B 2	32	1284	0.025			Line B 2	316	2188	0.144	
		Mean ratio		**0.087**	**1.00**			**Mean ratio**		**0.044**	**0.50**			**Mean ratio**		**0.158**	**1.81**

Table 4.10: Raw luminometer readings for transfected HEK293T cells measuring firefly (F) and Renilla (R) luciferase activity in relative light units (RLU).

Time Point	Test Replicate	Luciferase Activity (RLU)		Ratio (F/R)	Relative Fold Change
		Firefly (F)	Renilla (R)		
		pGL4.10			
24 hrs	1	6,936	2,840	2.442	
	2	6,516	2,120	3.074	
			Mean ratio	**2.758**	
48 hrs	1	11,096	29,872	0.371	
	2	9,924	21,016	0.472	
			Mean ratio	**0.422**	**1.00**
		pGL4/*IMPG2*			
24 hrs	1	2,868	2,148	1.335	
	2	2,780	2,248	1.237	
			Mean ratio	**1.286**	**0.47**
48 hrs	1	5,932	24,816	0.239	
	2	5,868	24,440	0.240	
			Mean ratio	**0.240**	**0.57**
		pGL4 mIL-17			
24 hrs	1	14,888	4,148	3.589	
	2	13,856	4,308	3.216	
			Mean ratio	**3.403**	**1.23**
48 hrs	1	15,572	29,620	0.526	
	2	17,428	37,484	0.465	
			Mean ratio	**0.495**	**1.17**

Table 4.11: Mean fold change calculated from firefly luciferase/Renilla luciferase ratios from RLU luminometer readings for canine skin fibroblast cell lines A and B assayed after each time point (24 hours, 48 hours and 72 hours) post-transfection, and one line of HEK293T cells assayed 24 hours and 48 hours post-transfection. Fold change differences are expressed relative to background control plasmid, pGL4.10 (fold change = 1). The standard deviation was calculated using mean ratios of each canine skin fibroblast cell line at each time point post-transfection (24 hours, 48 hours and 72 hours). The standard error was ascertained from the standard deviation with two incidences (cell line A and cell line B). Standard error values were plotted in fold change graphs in Figure 4.22 and Figure 4.23.

Time point post-transfection/cell line	**pGL4/*IMPG2*-promoter fold change relative to pGL4.10**	**pGL4-mIL17 fold change relative to pGL4.10**
	Canine skin fibroblast cells	
24 hrs		
Fibroblast line A	0.550	5.130
Fibroblast line B	0.820	4.900
Mean fold change	**0.690**	**5.010**
Standard deviation	0.186	0.162
Standard error	0.132	0.114
48 hrs		
Fibroblast line A	0.320	1.750
Fibroblast line B	0.530	2.750
Mean fold change	**0.430**	**2.250**
Standard deviation	0.148	0.708
Standard error	0.105	0.500
72 hrs		
Fibroblast line A	0.330	1.540
Fibroblast line B	0.500	1.810
Mean fold change	**0.420**	**1.670**
Standard deviation	0.124	0.192
Standard error	0.088	0.136
	HEK293T cells	
24 hrs	**0.470**	**1.230**
48hrs	**0.570**	**1.170**

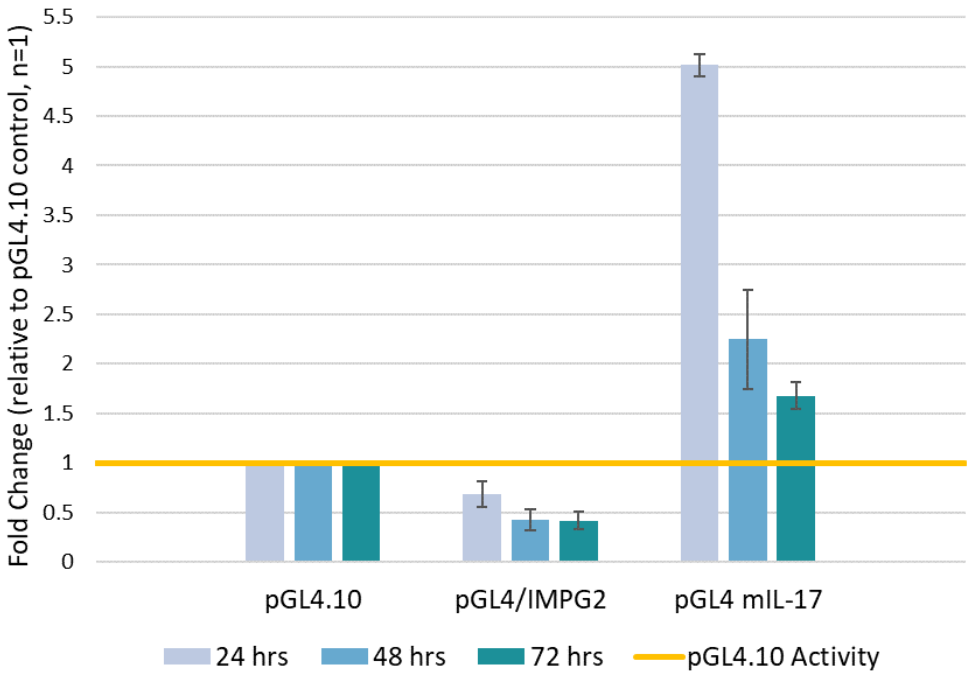


Figure 4.22: Mean fold changes following a luciferase assay in canine skin fibroblast cells of firefly luciferase/Renilla luciferase activity to assess the pGL4/*IMPG2*-promoter construct and pGL4 mIL-17 plasmid relative to a background control plasmid, pGL4.10. Data represent two independent experiments of two lines of canine skin fibroblast cells with triplicate or duplicate measures assessed at 24 hours, 48 hours and 72 hours post transfection. Error bars represent standard error. The yellow line represents the pGL4.10 (n = 1) baseline.

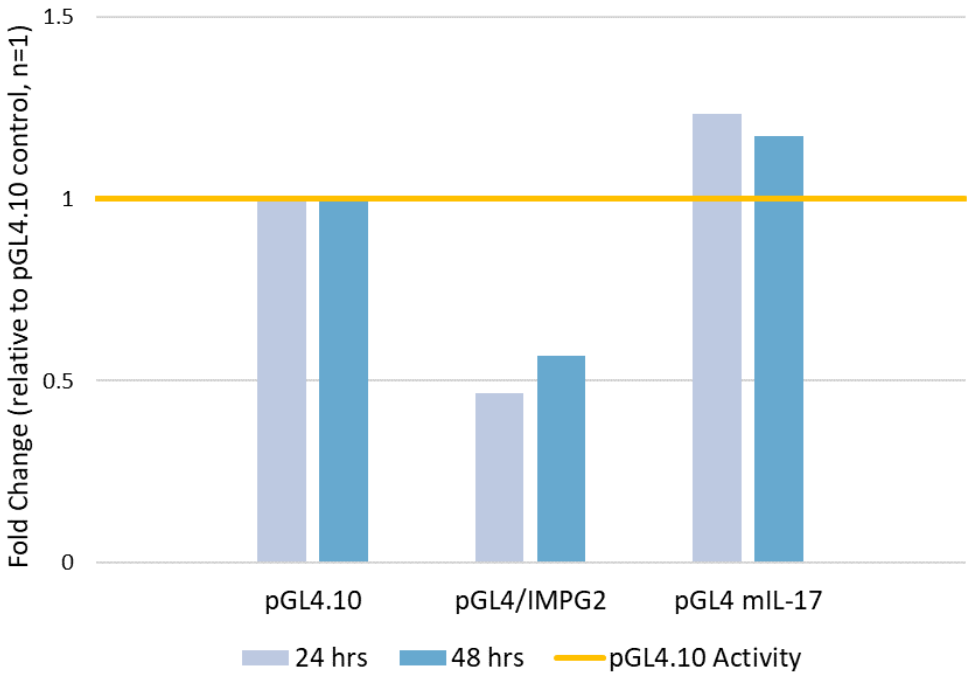


Figure 4.23: Mean fold changes following a luciferase assay in HEK293T cells of firefly luciferase/Renilla luciferase activity in pGL4/*IMPG2*-promoter construct and pGL4 mIL-17 plasmid relative to a background control plasmid pGL4.10 assessed at 24 hours and 48 post-transfections. The yellow line represents the pGL4.10 (n = 1) baseline.

4.3.8 DNA testing

Validation of the PRA association with the LINE-1 insertion enabled the development of a DNA test (termed 'PRA4') at the AHT to help reduce the incidence of this form of PRA in the LA. At the time of writing, 911 LA dogs from 22 countries (including the UK) have been genotyped for the *IMPG2* LINE-1 insertion with an allele frequency of 0.11. Genotyping data and allele frequencies are summarised in Table 4.12.

Table 4.12: Total number of LA dogs DNA tested for PRA4 showing genotypes and allele frequencies of the LINE-1 insertion. The wild type allele is represented by '+' and the mutant allele by '-'.

Cohort / Genotype*	*IMPG2* LINE-1 -/-	*IMPG2* LINE-1 +/-	*IMPG2* LINE-1 +/+	Total	Allele frequency
Across 22 countries	6	170	735	911	0.11
UK only	4	107	457	568	0.11

A LA dog identified by the DNA testing service at the AHT as homozygous for the *IMPG2* LINE-1 insertion underwent clinical follow-up at the AHT and was examined by a board-certified ophthalmologist/ BVA panellist. Upon ophthalmoscopic evaluation at the age of 2.5 years, the LA had early retinal abnormalities consistent with PRA, including tapetal hyperreflectivity, mild attenuation of blood vessels in the retina and changes to the optic disc colouration (Figure 4.24).

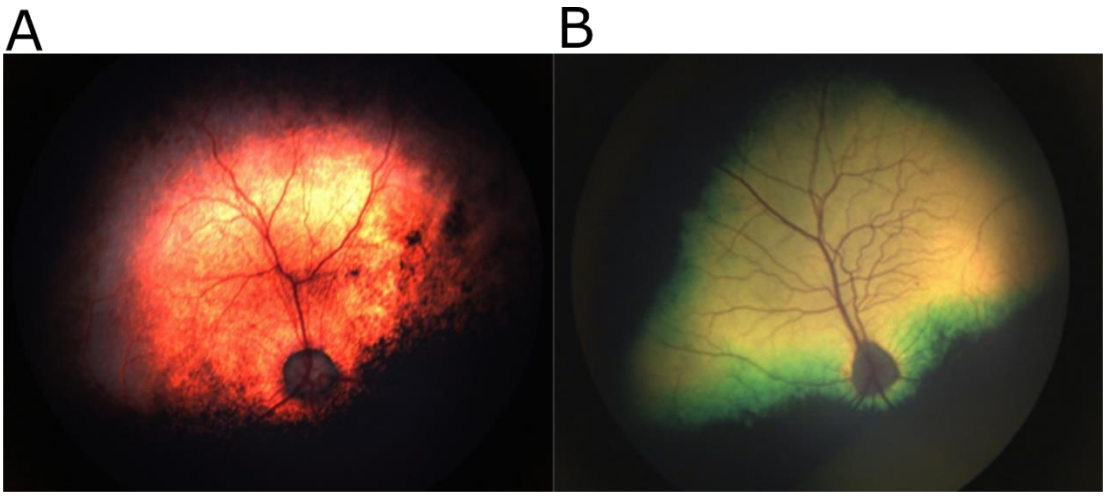


Figure 4.24: Fundus changes observed in a LA with early stage PRA. **(A)** An image of the retina in the left eye of a LA dog homozygous for the *IMPG2* LINE-1 insertion. Bilateral tapetal hyperreflectivity was observed as well as mild vascular attenuation and changes to the optic disc colouration. **(B)** An image of the retina of a control dog (GS dog) with a normal fundus (Hitti et al, 2019).

4.4 Discussion

In this study, a GWAS was performed to identify an interval associated with a novel autosomal recessive form of PRA in the LA. A statistically significant association was identified on CANFA33 that remained significant after correcting for multiple testing and population stratification. Analysis of a 1.3 Mb region of homozygosity on CANFA33, which was present in the GWAS PRA cases and absent in the controls, identified a LINE-1 insertion located within the predicted promoter region of *IMPG2.* PRA in the LA has not been reported in the literature; however, the disease is well recognised anecdotally in the breed and is listed on the BVA/KC/ISDS eye scheme in the UK. Pedigree analysis of individuals used in the present study indicated an autosomal recessive mode of inheritance.

The 1.3 Mb disease-associated region identified from GWAS analysis was homozygous in 12 of the 15 cases that passed QC. Two of the three cases that were not homozygous for the critical region were aged 6.3 years and 10 years, with a BVA certificate or veterinary referral letter diagnosing PRA, respectively. The third dog had been examined by a certified veterinary ophthalmologist, with a diagnosis of suspected PRA at 11 years of age, with additional clinical notes reporting that related dogs became blind due to a different eye condition: SARD. A fourth PRA case, submitted after the initial GWAS but included in variant follow-up, was also found to be homozygous for the wild type allele for all four variants of interest within the critical region, including the LINE-1 insertion. This case was diagnosed with PRA at the age of 6.8 years by a Member of the Royal College of Veterinary Surgeons (MRCVS) and was unable to visit a BVA panellist or a certified ophthalmologist to confirm the diagnosis. These four discordant cases are assumed to be affected with a genetically distinct form of PRA or a PRA phenocopy. A secondary GWAS analysis was conducted with three of the cases that were genotyped for the GWAS but were not homozygous for the critical region, using the remaining PRA non-affected LA from the GWAS as control dogs. This analysis revealed no suggestive loci. Recruitment of additional LA PRA cases that are clear of the *IMPG2* LINE-1 insertion may provide scope for future study of a second form of PRA in the breed.

Two retinal candidate genes, *IMPG2* and *CEP97*, are situated within the defined critical region on CANFA33. Both genes were manually interrogated in IGV software for potential causal variants using WES data generated from LA cases and controls in a previous study in our laboratory, which confirmed conclusions drawn from prior analysis of these WES data that no candidate exonic variants for PRA in this breed were found across the exome or within the defined critical region. This suggested that the PRA-associated variant was within a non-coding region of the genome not captured by the WES probes, including upstream promoter regions. WGS was therefore performed on one LA PRA case to provide a comprehensive genomic dataset. A PRA case homozygous for the critical region was chosen for WGS, to ensure it was representative of the

other cases from the GWAS sharing this haplotype. The critical region was explored and a LINE-1 insertion upstream of the *IMPG2* gene was identified. Notably, no strong exonic candidate variants were identified in *CEP97*; however, an intronic variant in *CEP97* was considered and genotyped in a LA cohort. Given the absence of recombination events between the LINE-1 insertion, the most associated GWAS SNP and the two intronic SNVs (in *CEP97* and *IMPG2*), the genotype frequencies were compared. Alleles illustrated in Figure 4.9 show that all four variants are in close proximity to one another, and indicates recombination events have occurred in two dogs between these regions. The LINE-1 insertion was considered the most plausible variant as it was a better functional candidate. Although the *IMPG2* intronic SNV is as correlated as the LINE-1 insertion, predicted pathogenicity and disruption of the *IMPG2* promoter region suggested the LINE-1 insertion as the most likely causal variant of PRA in these dogs.

Dog breeds exist as isolated populations each with a limited number of founders that has led to large regions throughout the genome in LD (Goldstein et al, 2006; Sutter et al, 2004). The significant LD that may be present in individual breeds means that it can be impossible to statistically refine the number of possible causal variants down to one, where regions of homozygosity and variants in LD with one another flank a disease locus. Studying additional individuals to continue to monitor genotype-phenotype concordance is important in these instances.

Mutations in *IMPG2* cause autosomal recessive RP (Bandah-Rozenfeld et al, 2010) and childhood-onset rod-cone dystrophy with early macular involvement in humans (Khan & Al Teneiji, 2019). Bandah-Rozenfeld et al. (2010) identified seven different mutations in patients with early-onset RP (five nonsense mutations and a 1.8 Kb genomic deletion over exon 9) and maculopathy (one missense mutation). IMPG2 belongs to a group of glycosylated proteins termed proteoglycans, which bind the large carbohydrates (glycosaminoglycans) in neural tissues. The retina consists of a neural network of layer-by-layer structures in which proteoglycans are secreted from photoreceptor cells and reside in the extracellular matrix bound to the RPE (Inatani & Tanihara, 2002). The IPM is a unique extracellular complex surrounding retinal photoreceptor outer segments and the RPE in the fundus of the eye, and is crucial for supporting normal function of retinal photoreceptors (Kuehn et al, 2001; Lazarus & Hageman, 1992). Studies have suggested that the IPM plays a role in recycling photoreceptor outer segments; in retina-RPE adhesion; the establishment of a milieu suitable for photoreceptor survival; and in the exchange of molecular products between the RPE and photoreceptor cells (Acharya et al, 2000; Kuehn & Hageman, 1999; Lazarus & Hageman, 1992). The role of *IMPG2* in retinal photoreceptors and its association with human retinal disease therefore makes it a strong candidate gene for canine retinal disorders.

Belonging to a group of transposable elements, LINE-1 elements are repetitive sequences present throughout mammalian genomes. The majority are inactive, defective elements that vary in size

(Ostertag & Kazazian, 2001). Structurally, they contain a 5' untranslated region (UTR) with internal promoter activity, two ORFs, a 3'-UTR and a poly-A tail (Ostertag & Kazazian, 2001). ORF1 and ORF2 are approximately 1 Kb and 4 Kb in size, respectively (Fanning & Singer, 1987). Agarose gel separation and NGS analysis demonstrated that the LINE-1 insertion in the predicted *IMPG2* promoter is a partial transposable element. Analysis of the LINE-1 sequence using *in silico* prediction tools confirmed this, where 5' truncation results in the total loss of ORF1 and partial loss of ORF2, with 1,629-bp aligning to OFR2 of the LINE-1 consensus sequence 'L1_cf'. There are a variety of mechanisms in which LINE-1 insertions can alter gene expression. These include: (i) altering gene splicing; (ii) introducing additional genes, regulatory regions, promoter sequence and/or alternative splice sites or start sites; and (iii) disruption of existing cis-regulatory elements (Belancio et al, 2009; Brooks et al, 2003; Credille et al, 2009; Feschotte, 2008; Ostertag & Kazazian, 2001; Smith et al, 2011), which in turn could insert or remove exons within genes, or alter promoter function and affect the expression of nearby genes.

Promoter regions are DNA sequences typically located upstream of a gene that, along with transcription factors interacting with the promoter region, determine where transcription is initiated. Transcription factors recognise short DNA sequences, called cis-regulatory sequences (between 100-500 bp in length), which consist of binding sites for sequence-specific transcription factors to determine which gene they will transcribe (Cherry et al, 2018). Promoter regions upstream of genes define the direction of transcription and indicate which DNA strand will be transcribed; therefore, mutations within promoter regions impact transcription regulation and are commonly associated with disease (de Vooght et al, 2009). In many eukaryotic genes a conserved TATA box promoter sequence (5'-TATAAA-3') is present and is recognised by transcription factors, thereby allowing RNA polymerase to bind and initiate transcription. Elongation begins with the addition of nucleotides as RNA polymerase moves along the DNA strand until it reaches a terminator to signal transcription to end. Chen et al. (2003) showed that regulatory elements excluding the TATA box were present within 100-bp upstream of the 5' end of *IMPG2* exon 1 and were 100% conserved in human and mouse. Five regulatory elements including pineal regulatory elements (PIREs) were located within this 100-bp upstream region and four copies of the PIREs were located between 400-bp and 1,000-bp upstream. Transcription is regulated through these additional regulatory elements (Chen et al, 2003). Although, sequence obtained through *in silico* prediction tools shows the presence of a TATA box within the predicted promoter region, transcription regulation of the canine *IMPG2* gene may mimic human and mice regulation using PIREs. Sequence in the *IMPG2* predicted promoter sequence, upstream of the 17-bp target site duplication of the LINE-1 insertion, contains a potential TATA box (5'-TATAAA-3') shortly followed by a run of A nucleotides: CATTATAAATTAGCTTAAATAAAAA, which could be involved in transcription regulation. Cone-rod homeobox (CRX), a cone-rod homeobox-containing transcription factor/otx-like homeobox protein, is the binding partner of PIRE, inducing

transactivation of a PIRE reporter construct (Li et al, 1998), and is expressed in retinal photoreceptor cells. In the present study, *CRX* is also one of the genes encoding transcription factors represented by the 'V$BCDF' matrix family in the *in silico* prediction tool 'MatInspector' (Cartharius et al, 2005). Thus, the presence of PIREs is likely to be important in controlling the expression of *IMPG2* in photoreceptors in the retina (Chen et al, 2003). Moreover, these TFBS elements may be disrupted by the insertion of the LINE-1 sequence in LA PRA cases, which in turn may impact IMPG2 transcription and protein function. An example of a LINE-1 element within a promoter region associated with disease was described by Davidson et al. (2016) in human patients with autosomal dominant corneal endothelial dystrophies. Four mutations within a conserved promoter region of the *OVOL2* gene were suggested to alter predicted TFBSs. Dysregulated *OVOL2* expression impacted the function of downstream genes and pathways, including transcriptional regulation.

Transposable elements located in non-exonic regions have been associated with inherited diseases in dogs and present compelling cases for causality of their respective conditions. In a study investigating an X-linked recessive bleeding disorder, haemophilia B, in German wirehaired pointers, an intronic LINE-1 insertion with a 200-bp poly-A tail was detected in affected dogs by Southern blot analyses (Brooks et al, 2003). Pedigree information enabled the LINE-1 insertion to be traced back through at least five previous generations and it segregated with the disease. An intronic LINE-1 insertion in the dystrophin gene was identified in Corgi dogs with Duchenne-like muscular dystrophy (Smith et al, 2011). This insertion introduced a new exon with a premature stop codon. More specifically, examples of non-exonic insertions of transposable elements have been associated with canine retinal diseases. These include a short interspersed nuclear element (SINE) insertion identified near the splice acceptor site of *FAM161A* in PRA-affected Tibetan Spaniel and Tibetan Terrier dogs (Downs & Mellersh, 2014); and a retinopathy in Swedish Vallhund dogs, where an intronic full LINE-1 insertion in a putative regulatory region of the *MERTK* gene was found to be associated with disease (Everson et al, 2017). In order to determine the direct impact of a transposable element on gene regulation or expression, the appropriate tissue (or blood if the gene is expressed in blood) from affected individuals is required. Further studies comparing retinal tissue between LA PRA cases and controls could shed light on the effect of the LINE-1 insertion on *IMPG2* expression: 5'RACE (Rapid Amplification of cDNA Ends) would confirm whether the insertion changes the 5'-UTR and its TFBSs; and RNA-seq analyses could reveal *IMPG2* expression and transcription regulation differences. As suitable tissues were not available, these experiments could not be undertaken for this study.

As no retinal (or *CRX* expressing) tissue was available from any cases in the current study, a luciferase assay was attempted using canine skin fibroblast cells, chosen because cell lines were available for immediate use. Expression of the *IMPG2* gene was confirmed by qPCR prior to

commencing plasmid construction and cell transfection for a reporter assay. Despite the observation that *IMPG2* was expressed in canine skin fibroblast cells, the poor luciferase expression observed using a luciferase reporter assay suggests that the *IMPG2* promoter inserted in the pGL4/*IMPG2*-promoter plasmid did not act as a promoter in the canine skin fibroblast cells, or the HEK293T cells. However, the use of a better positive control with higher RLU readings would confirm this assumption. Contrary to initial assumptions, the pGL4 mIL-17 plasmid was not a suitable positive control for the experiment as the mouse IL-17 promoter did not prove to be active at a high enough level in canine skin fibroblast cells or HEK293T cells. IL-17 is an inflammatory cytokine; therefore, the promoter may not be active in these cell lines under the normal culture conditions used. An appropriate positive control should result in RLU values and relative fold changes much higher than those obtained in the current study, comparable to previous IL17-promoter-driven luciferase assays (Chuang et al, 2018; Singh et al, 2018), and should be robustly expressed in the cell types under observation. Repeating the transfections with a more appropriate positive control containing a promoter that is constitutively expressed at high levels in most cell types, such as a cytomegalovirus (CMV) or a SV40 driven promoter, would confirm this hypothesis. The GFP reporter gene transfections assessed transfection efficiency in the current study and confirmed that a CAG promoter with a CMV enhancer is active in the canine skin fibroblast skin cells, as well as the HEK293T cells. Therefore, a plasmid constructed with a CMV promoter/enhancer could be explored as a more suitable positive control for future studies. In addition, reasonable levels of *hRluc* were detected in canine skin fibroblast skin and HEK293T cells, indicating a vector with a SV40 promoter is active in both cell types. Given that *IMPG2* is expressed at a low level in the canine skin fibroblast cells, and that transfection efficiencies in these cell lines were relatively low in comparison to HEK293T cells, these factors could also have contributed to the low values obtained from the luciferase assay. Nevertheless, the hypothesised absence of photoreceptor-specific-TFBSs or *CRX* expression in canine skin fibroblast cells and HEK293T cells are probable confounding factors to an unsuccessful luciferase reporter assay. Although *IMPG2* expression was ascertained in the canine skin fibroblast cells, albeit at a lower level than in retinal tissue, it appears that photoreceptor-specific-TFBSs such as CRX are crucial for normal *IMPG2* promoter activity. Thus, a luciferase reporter assay using retinal cell lines would be required to measure *IMPG2* promoter activity and the impact of the identified LINE-1 mutation on *IMPG2* expression. As no retinal cell lines were available to use in a luciferase reporter assay, a second assay with a more appropriate positive control was not pursued. Since the hypothesis that transcription was disrupted by the LINE-1 element could not be tested either by luciferase assay or RNA or protein expression analysis, the effect of the LINE-1 insertion observed in this study on the *IMPG2* gene can only be speculated.

DNA tests are used by dog owners and breeders as a tool to avoid producing puppies with a particular inherited condition. One outcome of this study has been the development of a DNA test

at the AHT, termed 'PRA4', to enable LA dogs to be tested for this form of PRA. In the UK LA population, tested across a two year duration since the DNA test was made available, the recently estimated mutant allele frequency of 0.11 indicates that one in 100 LA dogs are expected to be affected with this form of PRA, and 18 in 100 are expected to carry the mutation. Although this population will probably include closely related dogs, this value is well within the range presented by estimated allele frequencies of other recessive conditions in canine studies (Downs et al, 2011; Downs et al, 2013; Zangerl et al, 2006).

Clinical follow-up of one *IMPG2* LINE-1 homozygote provides some evidence that the age of onset in the breed may be variable, where clinical signs of retinal changes can be present from as early as 2.5 years of age; although examination of additional dogs is required to further characterise an age of onset. Where provided with sample clinical information, the age of diagnosis of PRA cases homozygous for the disease-associated haplotype from the GWAS varied, ranging from aged 1.75-10 years. The owner of the LA dog that was identified through the DNA testing service as homozygous for the *IMPG2* LINE-1 insertion had not noticed behavioural changes or signs that the dog's vision was deteriorating; however, early signs PRA were apparent upon ophthalmoscopic examination. Dogs are remarkably good at adapting well to blindness, especially when vision loss is gradual. Factors such as the presence of other canine companions in the household and a consistent environment can also aid in this adaptation. Continual annual checks of this PRA-affected LA and other genetically affected LA dogs will help describe the rate of progression of PRA in this breed.

4.5 Conclusion

Using GWAS and WGS, as well as an existing WES dataset, a LINE-1 insertion upstream of the *IMPG2* gene that strongly segregates with PRA in LA dogs was identified. Due to the lack of tissue from affected dogs or suitable retinal cell lines, no functional analysis or a luciferase reporter assay could be conducted to determine the impact of the LINE-1 insertion on the *IMPG2* gene. However, literature surrounding *IMPG2* and its implication in human retinal disease, alongside previous studies of transposable elements in non-exonic or gene promoter regions in inherited diseases, strongly suggests that the LINE-1 insertion is a causal PRA variant in the LA dogs under investigation. Extensive genotyping of this variant in multiple breeds strongly implied that the LINE-1 insertion is private to the LA breed, or at least rare amongst other breeds, and was only present in PRA-affected LA dogs. Elucidation of this PRA-associated variant facilitated the development of a DNA test that, over time, will help reduce the frequency and incidence of this mutation in the LA breed worldwide.

Chapter 5 PRA in the Shetland Sheepdog

5.1 Introduction

The Shetland Sheepdog (SS), also referred to as the Sheltie, is a popular breed in the UK. In 2018, it had the third highest registrations with The Kennel Club, UK within the Pastoral group, with 794 registrations. Population analysis by The Kennel Club, UK of the SS breed between 1980 and 2014 shows a loss of genetic diversity has occurred, with higher rates of inbreeding due to the use of popular sires and a reduced estimated effective population size of 77.6 (a value below 100 indicates an increase in the rate of loss of genetic diversity within the breed) (The Kennel Club UK, 2015). Therefore, despite a high number of yearly registrations, many of these SS may be closely related and today's SS population may continue this trend of having high coefficient of inbreeding rates and a smaller gene pool. This increases the probability that two copies of the same variant in a gene being inherited from an ancestor common to both the sire and dam in their offspring.

Shelties are known to suffer from two main hereditary ocular conditions: collie eye anomaly (CEA) (Parker et al, 2007) and PRA (Wiik et al, 2015). One form of PRA has previously been described in the SS, caused by an exonic 4-bp deletion in the cyclic nucleotide gated channel alpha 1 (*CNGA1)* gene, however genetic heterogeneity was indicated in the study cohort (Wiik et al, 2015). Several SS that were unrelated to the *CNGA1* affected proband were clear of the deletion, suggesting that at least one additional form of PRA was segregating in the breed. Prior to these findings, the only form of inherited retinal disease for which a disease-associated variant has been reported in the SS is CEA (Parker et al, 2007). As of 2018, PRA is no longer listed by the BVA/KC/ISDS eye scheme due to no recent cases reported under the scheme (Appendix 3). However, at the time this study was initiated, PRA was of concern to the English SS Breed Club in the UK as at least one non-*CNGA1* PRA case was known. Knowledge that non-*CNGA1* forms of PRA are present in the breed, with an unknown number of carriers in worldwide populations, warranted further study of PRA to identify genetically unresolved incidences. The aim of this study was to identify the genetic cause of an additional form of PRA, enabling DNA test development for breeders to use as a tool to reduce the frequency and eventually eradicate a novel form of PRA from the breed.

5.2 Materials and methods

5.2.1 PRA diagnosis and sample collection

PRA in the SS is clinically indistinguishable from PRA in other breeds of dog and is diagnosed based on ophthalmoscopic evaluation by a board-certified veterinary ophthalmologist or a qualified BVA/KC/ISDS eye scheme panellist (or the European equivalent if outside the UK). Buccal mucosal cheek swabs were submitted to the AHT from four SS PRA cases, one suspected PRA case, and 89 SS PRA controls. Of the 89 controls, 43 SS were aged eight years and over and had a clinically clear PRA diagnosis following ophthalmic examination by a veterinary ophthalmologist via the BVA/KC/ISDS Eye Scheme in the UK, or the European equivalent(s), and were therefore considered to be PRA controls for this study. Of the four PRA cases, two were littermates. This litter comprised three SS with one female and one male SS clinically affected with PRA (one diagnosed by a BVA/KC/ISDS eye panellist; one diagnosed by a board-certified veterinary ophthalmologist), and one male suspected to have PRA.

The age of diagnosis for the four cases ranged from 1.26 - 8.93 years of age. Where known, owners of SS dogs included in the study resided in the UK (*n* = 73), Jersey (*n* = 1), The Netherlands (*n* = 3) and Czech Republic (*n* = 1). DNA was extracted from buccal mucosal cheek swabs using protocols described in Chapter 2 (section 2.7).

5.2.2 Exclusion of known retinal mutations

Twenty-six previously published mutations, including the *CNGA1* variant identified in the SS breed (Wiik et al, 2015), were excluded as mutations responsible for this form of PRA in the SS (see Chapter 2, section 2.7).

5.2.3 Whole genome sequencing (WGS)

gDNA from a PRA-affected SS was prepared for WGS. WGS was performed using protocols previously described (see Chapter 2, section 2.6). gVCF files from 186 canine WGS in the AHT dataset were merged into a multi-sample VCF file.

5.2.4 Variant filtering

Variants were filtered for those with the highest effect scores (4 and 5) (see section 2.8 for effect score grading) and for variants segregating as an autosomal recessive disease. An X-linked mode of inheritance was also considered by filtering for high impact variants located on the X chromosome and by checking for XLPRA mutations previously identified in other breeds that could not be excluded prior to WGS. Sequencing reads from WGS data were visualised in IGV to determine genotypes of the known XLPRA-associated mutations and a *de novo* assembly of sequencing reads across regions which were unannotated in the reference genome, as described in Chapter 2 (section 2.6). In addition, the genome data of six additional genes associated with X-

linked retinal diseases in humans: *RP2* (Bhattacharya et al, 1984; Hardcastle et al, 1999), *OFD1* (Coene et al, 2009; Hardcastle et al, 2000; Webb et al, 2012), *TIMM8A* (Jin et al, 1996), *CACNA1F* (Bech-Hansen et al, 1998; Strom et al, 1998) and *NYX* (Bech-Hansen et al, 2000; Gal et al, 1989; Hardcastle et al, 1997) were visualised for variants.

Further variant filtering was conducted using the DBVDC (see Chapter 2 section 2.8) as a genome bank. *De novo* local realignments of sequencing reads were conducted over each filtered variant region using HaplotypeCaller (Poplin et al, 2018).

5.2.5 *In silico* protein prediction tools

To assess the impact of the candidate variant amino acid substitution on gene protein function, bioinformatics tools Mutation Analyzer, PolyPhen-2, and SIFT were used to computationally predict the effect of the missense variant.

5.2.6 Comparative species conservation

The Evolutionary Conserved Region Browser (http://ecrbrowser.dcode.org) and UCSC Genome Browser were used to determine the degree of sequence conservation of the *BBS2* gene across species. Using the human genome (GRCh37/hg19) as the base genome, *BBS2* human coordinates were compared across five species.

5.2.7 Screening Shetland Sheepdogs

A cohort of SS controls that were aged eight years or over and diagnosed as clinically clear through the BVA/KC/ISDS eye scheme were genotyped for variants highlighted through the variant filtering process. Primers were designed around each variant to generate PCR amplicons larger than 300-bp in size for Illumina short-read sequencing on a MiSeq platform (Illumina). Primers are listed in Appendix 4. A multiplex PCR was conducted as described in Chapter 2 (section 2.7) followed by PCR purification, library preparation and data interpretation as described in section 2.6. The University of Helsinki, Finland, kindly genotyped a cohort of 394 SS dogs including one PRA case, two dogs with suspected PRA and two other retinopathy-affected dogs using the allelic discrimination assay described below. The Norwegian University of Life Sciences, Norway, also kindly sequenced eight PRA cases SS for the *BBS2* variant using primers and conditions described below.

5.2.8 Variant validation and genotyping

The missense *BBS2* SNV was confirmed using Sanger sequencing (see section 2.5). An allelic discrimination probe-based assay was designed to effectively distinguish between individuals homozygous for the SNV (C/C), heterozygous (G/C) or homozygous for the wild type/ common allele (G/G). Allelic discrimination assays were performed as described in Chapter 2 (section 2.9) using a 40X primer/probe mix and conditions listed in Appendix 4. The *CEP290* SNV was

genotyped using PCR followed by Sanger sequencing as described in Chapter 2 (sections 2.3, 2.4.1 and 2.5). Primers and assay conditions are listed in Appendix 4.

5.2.9 Extracting all variants in BBS-genes

To highlight additional variants in the SS WGS dataset that were situated in other BBS-associated genes, all variants (regardless of their effect score) were extracted from the multi-VCF file into a text file. The following 'grep' Unix command was used to extract all variants that were situated in 21 known BBS-genes, including gene synonyms in the search:

```
grep -wE
"(Chr|BBS1|BBS2|ARL6|BBS4|BBS5|MKKS|BBS7|BBS8|TTC8|BBS9|BBS10|TRIM32|BBS12|MKS
1|CEP290|WDPCP|SDCCA8|LZTFL1|BBIP1|IFT27|IFT172|C8orf37|BBIP10)" input_file.txt >
output_file.txt
```

5.3 Results

5.3.1 Clinical signs

PRA diagnosis

Pedigree information revealed that the SS elected for WGS belonged to a litter consisting of three offspring: one female and two males (male A and male B). A PRA diagnosis was confirmed in the female SS and one of the male SS dogs (male A) by a BVA/KC/ISDS panellist and/or veterinary ophthalmologist. The female SS was 8.7 years at the time the PRA diagnosis was confirmed. The same dog had been examined under the BVA/KC/ISDS eye scheme 16 months earlier (aged 7.4 years), by the same ophthalmologist, and was clinically unaffected at that time; indicating a form of PRA with a relatively late age of onset. PRA was diagnosed based on findings of bilateral retinal degeneration due to bilateral tapetal hyperreflectivity with retinal vascular attenuation and an optic disc pale in colour. Menace and pupillary light responses were present but reduced. No cataracts were present. Additional findings were bilateral distichiasis, bilateral nuclear sclerosis and asteroid hyalosis. Night blindness had been apparent for four weeks prior to the confirmed diagnosis and a noticeable decline in daylight vision followed thereafter. A maze test was performed, where the female case collided with obstacles under scotopic conditions but avoided obstacles under photopic conditions. An ERG showed depressed rod and cone response. Ophthalmoscopic follow-up examination at 11.3 years reported no PRA progression, but cataract formation at the equatorial region. Mature cataracts were present bilaterally by the age of 12 years.

Veterinary records show that male A was also diagnosed with retinal abnormalities that were consistent with PRA by a board-certified veterinary ophthalmologist. This included tapetal hyperreflectivity and bilateral retinal blood vessel attenuation at the age of six years. Although male B was not formally examined through the BVA/KC/ISDS eye scheme or by a board-certified veterinary ophthalmologist, a PRA diagnosis was suspected. The owners of male B reported the dog had started to suffer from visual impairment around the age of 9-10 years, initially presenting with reduced vision in low light or at night that progressed until his death at the age of 15 years. The dam of the litter was examined under the BVA/KC/ISDS scheme at 11.4 years and was clinically unaffected at that age. There is no evidence that the sire's eyes were ever examined by a veterinary ophthalmologist.

For the remaining two confirmed PRA cases, one was diagnosed with early-onset PRA at the age of 1.26 years and one was diagnosed with PRA at the age of 8.93 years; both by separate BVA/KC/ISDS panellists.

Weight

Veterinary history regarding weight was only available from the age of 11 years for the female SS. Weight was recorded from the age of 11 years up until euthanasia at the age of 13.9 years, ranging from 5.1 kg – 6.6 kg with the dog on a low-fat diet.

Male A was obese, weighing 19.5 kg at the age of 5.5 years. The owner reported that the SS had a severely increased appetite and described the dog as 'ravenous' when offered food. A low-fat diet eventually reduced the dog's weight to 14 kg at the age of 6.4 years, but this was not maintained and resulted in subsequent weight gain. A dirlotapide based drug (Slentrol®) was used from the age of 5.9 years in an attempt to reduce body weight. No information on male B's weight was available, nor the two remaining PRA cases not in the litter.

Renal function

Limited information was available on renal biochemistry for the female SS case, however nephroliths were present at 13.3 years of age, and eventually the owner elected for euthanasia at 13.9 years due to renal failure. In male A, urea and creatine levels were within normal limits at 5.5 years of age: 7.9 moles per Litre (mol/L) (normal range 2.5- 9.6 mol/L) and 85 mol/L (normal range 44-159 mol/L), respectively. When assessed at eight years, these levels had increased (urea 26.7 mol/L; creatine 160 mol/L). Two months later, urea and creatine levels had significantly increased to 35.6 mol/L and 258 mol/L; resulting in a diagnosis of renal failure. The veterinarian prescribed a renal diet feed and Fortekor® to aid in the management of chronic renal failure. By the age of 9.4 years, kidney disease had progressed with elevated urea and creatine levels (urea 46.4 mol/L; creatine 444 mol/L) and euthanasia was elected. No information on male B's renal health was available, nor the two remaining PRA cases not in the litter.

Additional features

The female SS PRA case showed uncharacteristic features for the breed, including an upturned nose, an unusual coat texture and dental abnormalities (Figure 5.1). At the time, the condition was deemed purely cosmetic and no further follow-up was carried out. The owner of this dog shared speculations that one of the male dogs in the litter (male B) shared the visible uncharacteristic features seen in the female (including an upturned nose and unusual coat texture), with the female's distinctive nose shape more pronounced. These features became more noticeable as the dog developed, becoming more apparent around the age of six months. In addition to PRA, the female SS was diagnosed with distichiasis and struvite urolithiasis had previously been detected, which was controlled by an appropriate diet. Veterinary history also noted dental remarks where multiple teeth were extracted at the age of 11.9 years. Photographs confirm this female had abnormal teeth (Figure 5.1).

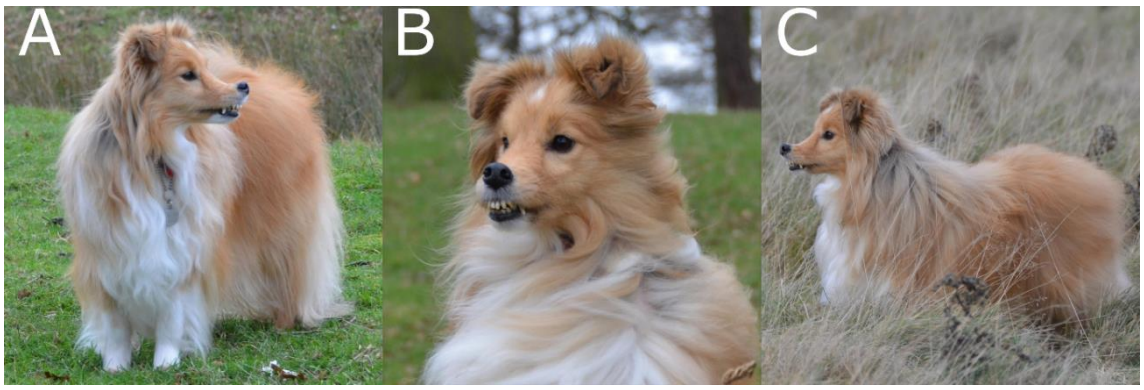


Figure 5.1: Photographs of the PRA-affected female SS in the initial litter show the uncharacteristic wavy coat, upturned nose and dental defects (**A-C**).

The only evidence obtained for the physical appearance of male B was the anecdotal evidence that it shared the distinctive characteristics of the female. No further information on the physical appearance of male A was available, although the dog was reported to have undergone dental treatment of three teeth extractions due to resorptive lesions. Photographs of both sire and dam show that they did not exhibit the unusual nose shape or coat texture observed in the affected offspring (Figure 5.2). Of the two PRA cases not in the litter, the 1.26-year-old SS showed normal physical characteristics typical for the breed, and no information on the physical appearance of the 8.93-year-old PRA case was available.

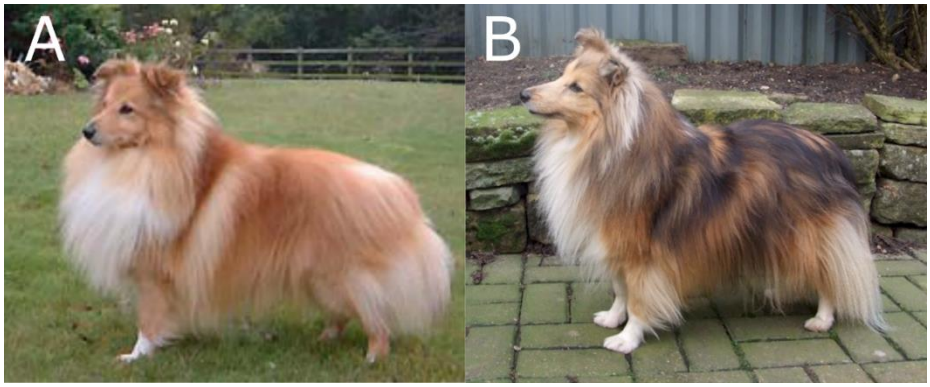


Figure 5.2: Photographs of the sire and dam to the PRA-affected SS. Both the sire **(A)** and dam **(B)** of the original litter of SS in the present study have typical physical characteristics and features that adhere to the SS breed standards for nose shape and coat texture.

5.3.2 Exclusion of known PRA-associated mutations

Twenty-seven previously published retinal mutations were excluded as the cause of PRA in the PRA-affected SS case, in addition to both the *NECAP1* and *IMPG2* variants identified in this thesis. This indicated that a novel PRA-associated mutation was present. Full materials and methods are described in Chapter 2 (section 2.7).

5.3.3 Variant filtering

At the time of study, 185 canine genomes that were sequenced as part of other studies at the AHT were available to use as non-breed matched controls to compare against the PRA-affected SS genome. Of these 185 genomes, 176 were of dogs that were recruited for non-PRA studies and were considered to be PRA controls for the purposes of this study. These comprised 87 pure breed dogs and 5 cross breed dogs. Date of births were known for 108 of the 176 non-breed matched controls, with dogs ages ranging from 0.11 and 15.4 years at the time of DNA submission, with a median age of 6.79 years (interquartile range 1.90-9.14). These dogs were affected with conditions including epilepsy, sensory neuropathy, hereditary cataracts, glaucoma, Spongiform Leukoencephalomyelopathy (SLEM) and primary lens luxation. The DBVDC contained 648 genomes at the time of study.

The merged VCF file contained 186 genomes (1 case versus 185 controls) with a total of 27,611,812 variants amongst all genomes. Filtering for variants with a high effect score (4 and 5) retained 136,383 high impact variants. As the litter consisted of one PRA-affected female and at least one PRA-affected male, an X-linked dominant mode of inheritance was considered. However, as the sire was apparently unaffected and the dam was clinically confirmed as unaffected, an X-linked dominant PRA was unlikely. Nevertheless, IGV software was used to visualise sequencing reads from the SS WGS over the entire *RPGR* gene, which has previously been implicated in canine XLPRA. No variants of interest were present in *RPGR*. In addition, no high impact variants segregated appropriately for an X-linked mode of inheritance (dominant or recessive) or were present in six genes associated with X-linked RP in humans: *RP2*, *OFD1*, *TIMM8A*, *CACNA1F* and *NYX*. Thus, an X-linked model was excluded, and an autosomal recessive mode of inheritance was considered most likely.

Filtering based on an autosomal recessive model, i.e. retaining high effect variants (effect score of 4 or 5) that were homozygous in the SS case and either heterozygous or homozygous for the alternate allele in all 176 non-breed matched controls, resulted in 151 variants. Further exclusion of variants located on unknown chromosomes narrowed the variant list down to 134 variants. These 134 variants were checked across canine genomes in the DBVDC, which comprised of 648 genomes at the time of study (32 belonging to the AHT), including three SS genomes. Twelve variants were all homozygous in the SS case and absent from all other genomes in the DBVDC. Five of these fifteen variants overlap the variants in the final filtering list, described below.

In parallel with screening the DBVDC, all genes in which the 151 variants were located were assessed for their candidacy: comprising 74 annotated genes (80 variants) and 40 Ensembl canine stable identifiers (ENSCAF IDs; 71 variants). Of the 40 ENSCAF IDs, human homologues could be determined for 10 genes (28 variants). Fifty-seven variants located within ENSCAF IDs of which no human orthologous region was present and/or those on unknown chromosomes were

excluded. Therefore, a total of 84 gene names (94 variants) were investigated in the OMIM database against a list of relevant keywords considered to be associated with PRA phenotypes (see Chapter 2, section 2.8). Thirty-eight variants were predicted to be tolerated by SIFT and 36 variants were homozygous in non-SS or non-Collie breed control genomes in the DBVDC, and were therefore excluded. Two of the remaining 20 variants were homozygous in SS genomes or Collie breeds. Manual visualisation of these 20 filtered variants in IGV following *de novo* local realignment using HaplotypeCaller (Poplin et al, 2018) confirmed that 15 variants were called correctly (Table 5.1). Of these 15 variants, two were highlighted as candidate variants based on direct, or indirect, relation to one or more of the PRA-related terms in the keyword list. Nevertheless, all 15 variants were investigated further.

5.3.4 Variant mining

The 15 variants identified in the variant filtering process were genotyped (by sequencing on the Illumina MiSeq) in a cohort of 43 SS control dogs, which were aged eight years or older and diagnosed as clinically clear of PRA. PCR amplicons could only be generated for 12 variants, of which 11 were homozygous in at least one of the SS control dogs and therefore were excluded as causal PRA variants. Table 5.1 summarises the 15 variants and determines whether they could be excluded as candidate variants or not. One candidate variant not excluded is a non-synonymous missense SNV in the Bardet-Biedl syndrome 2 gene (*BBS2*) (Figure 5.3); a gene previously associated with a ciliopathy termed BBS in humans.

Sequencing of three variants failed due to difficulty in amplification, probably due to repetitive sequence and/or GC-rich sequence. These were variants in the genes *GGN*, *C7* and *IRS2*. RNA-seq data generated for an unrelated study from a PRA non-affected Petit Bassett Griffon Vendéen canine retina showed the *GGN* gene is considered to be expressed at low levels in the retina, whereas *C7* and *IRS2* are considered to be highly expressed. Although genotyping by sequencing of the *GGN*, *C7* and *IRS2* variants failed, the variants were identified in additional breeds in the DBVDC genomes. The *GGN* 9-bp insertion was heterozygous in an Alaskan Malamute, two Irish Terrier dogs and a Wolf, and homozygous in a Chow Chow dog and two Wolf genomes. The *C7* SNV was present in two SS genomes in the DBVDC: one homozygote and one heterozygote. The DBVDC member that owned the SS genome homozygous for the *C7* variant confirmed the dog was diagnosed with a neurological condition by a neurologist and no ocular phenotypes had been reported. The *C7* variant was also heterozygous in WGS from a Lancashire Heeler dog included in the AHT's control genomes used in the analysis. Finally, the *IRS2* 3-bp insertion had poor sequencing coverage in the original SS WGS due to a high GC content across this region. When screened for in the DBVDC, no sequencing data were available for the *IRS2* variant in multiple genomes and two Dachshund dogs were heterozygous for the insertion. Although these variants

could not be fully excluded using PCR methods, the candidacy of *BBS2* SNV warranted prioritisation for further exploration.

Table 5.1: Fifteen variants identified through WGS were genotyped in a SS cohort. Filtering variants from WGS analysis identified 15 variants to follow up in additional SS. Two variants were highlighted as candidate variants based direct, or indirect, relation to one or more of the PRA-related terms in the keyword list, represented by a † symbol. However, only one gene has been associated with a retinal phenotype in any other species: *BBS2* (highlighted). Variants were sequenced in 43 SS controls aged eight years or older and were excluded as variants of interest if homozygous in at least one control SS. Variants with low confidence in sequencing reads due to low coverage or those where control dogs were only heterozygous or homozygous for the alternate allele were not excluded as potential candidate variants.

CanFam3.1 Chromosomal Position	Gene/Ensembl Stable Identifier	Variant	Excluded/Not Excluded
1: 110,147,994	CD3EAP	Missense SNV	Excluded
1: 116,038,553	ENSCAFG00000028805 (no human orthologue)	Missense SNV	Excluded
1: 119,547,008	ZNF507	1-bp deletion	Excluded
2: 62,484,625	CHD9	Missense SNV	Excluded
15: 52,265,347	FGG	Splice site insertion of transposable element	Excluded
16: 36,561,913	DLC1	Missense SNV	Excluded
16: 44,477,295	F11	Splice site SNV	Excluded
30: 24,115,675	ENSCAFG00000016650 (human orthologue PSMA1)	Missense SNV	Excluded
36: 10,135,919	SLC38A11	Missense SNV	Excluded
X: 1,762,952	MXRA5	Missense SNV	Excluded
X: 43,769,675	BMP15	Missense SNV	Excluded
1: 114,617,154	GGN	9-bp insertion	Not excluded
2: 59,693,737	BBS2†	Missense SNV	Not excluded
4: 68,451,901	C7	Nonsense SNV	Not excluded
22: 58,226,397	IRS2†	3-bp insertion	Not excluded

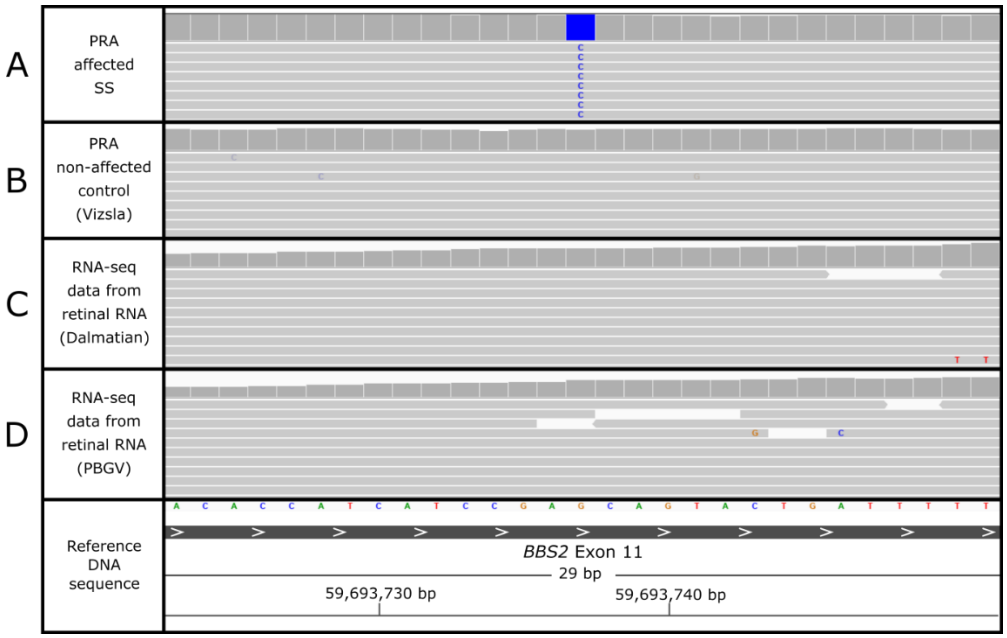


Figure 5.3: IGV display of the *BBS2* c.1222G>C SNV in the SS. **(A)** WGS from a SS PRA case homozygous for the *BBS2* SNV. **(B)** WGS from a PRA non-affected control genome (Vizsla dog) is homozygous for the wild type allele. *BBS2* is expressed in retinal tissue as demonstrated by RNA-seq alignment from two control dogs: **(C)** one Dalmatian dog and **(D)** one Petit Basset Griffon Vendéen (PBGV) dog, both of which are also homozygous for the wild type allele at this position.

5.3.5 Variant identification

Sanger sequencing confirmed the c.1222G>C missense SNV in exon 11 of 17 of *BBS2* (Figure 5.4). This non-synonymous variant results in an amino acid change from alanine to proline at amino acid position 408 (p.A408P). Unlike the canine *BBS2* gene located on the forward strand in the CanFam3.1 canine genome annotation, the human *BBS2* gene is located on the reverse strand. At the homologous position in the human *BBS2* gene (GRCh38; chr16:56,501,010), a missense variant (G/A) resulting in an alanine to valine amino acid change, has been reported (rs1274202413) with a minor allele frequency of < 0.01. No phenotype data are available for this variant, which is also absent from the Genome Aggregation Database (gnomAD), nor are any other variants at this position in gnomAD. Moreover, alanine and valine are structurally similar amino acid residues, therefore the substitution described above may not have a functional effect on the human BBS2 protein. Proline, however, is structurally different to alanine and is usually located in unstructured regions between secondary structures, and therefore it is important in shaping the secondary structure of a protein (Betts & Russell, 2007). This provides support that the alanine to proline substitution reported in SS in this study impacts canine BBS2 structure and function.

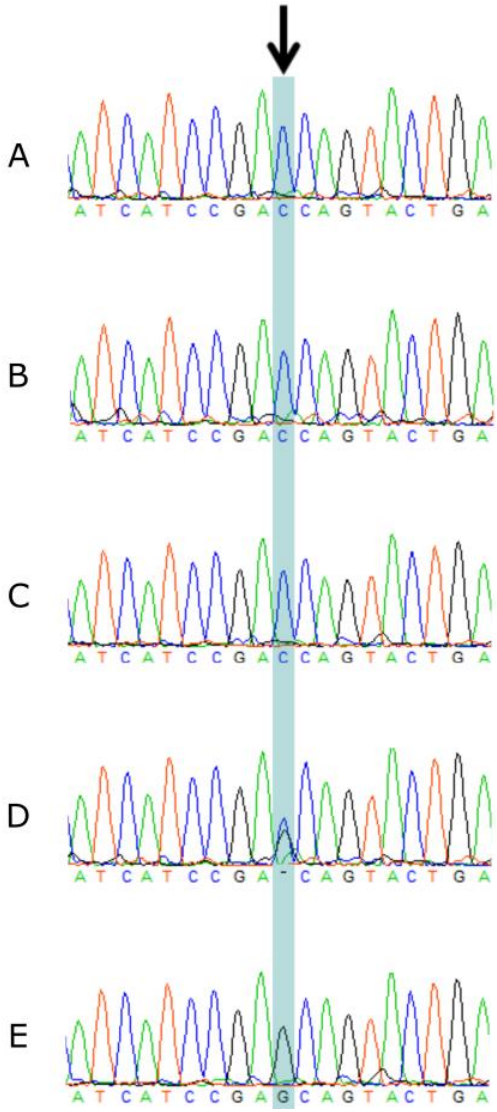


Figure 5.4: Sanger sequencing of the *BBS2* c.1222G>C variant confirmed the presence of the identified SNV in the SS DNA used for whole genome sequencing **(A)**. Two additional dogs (shown in Figure 5.10) were also homozygous for the c.1222G>C variant **(B, C)**. SS dogs heterozygous for the variant and homozygous for the wild type allele are shown in tracks **D** and **E**, respectively.

5.3.6 *In silico* protein predictions

Mutation Analyzer and PolyPhen-2 predicted the c.1222G>C; p.A408P amino acid change in *BBS2* to be rarely substituted and probably damaging to the protein (PolyPhen score 0.993), respectively. SIFT predicted the change to be deleterious with a SIFT score of 0.03.

5.3.7 Comparative species conservation

BBS2 is conserved across species showing 90% sequence similarity to the human orthologue. Additionally, the 'GCA' codon implicated is well conserved across 37 mammals, including humans. Using the human genome (GRCh37/hg19 assembly) as the base genome, the degree of evolutionary sequence conservation of *BBS2* was assessed in the zebrafish, chicken, mouse, dog (CanFam2.0 assembly), and rhesus macaque genomes (Figure 5.5A). The majority of coding exons

were conserved across each species, with the zebrafish and chicken genomes displaying a reduction in conservation across all exons and an absence of exons (exons 1, 3, 9, 11 and 12 in the zebrafish and exon 1 in the chicken). The rhesus macaque and the dog genomes share the highest degree of *BBS2* sequence conservation across the species evaluated, closely followed by the mouse genome. BBS2 amino acid residues are conserved across this region; specifically, the affected amino acid alanine is highly conserved through evolution (Figure 5.5B).

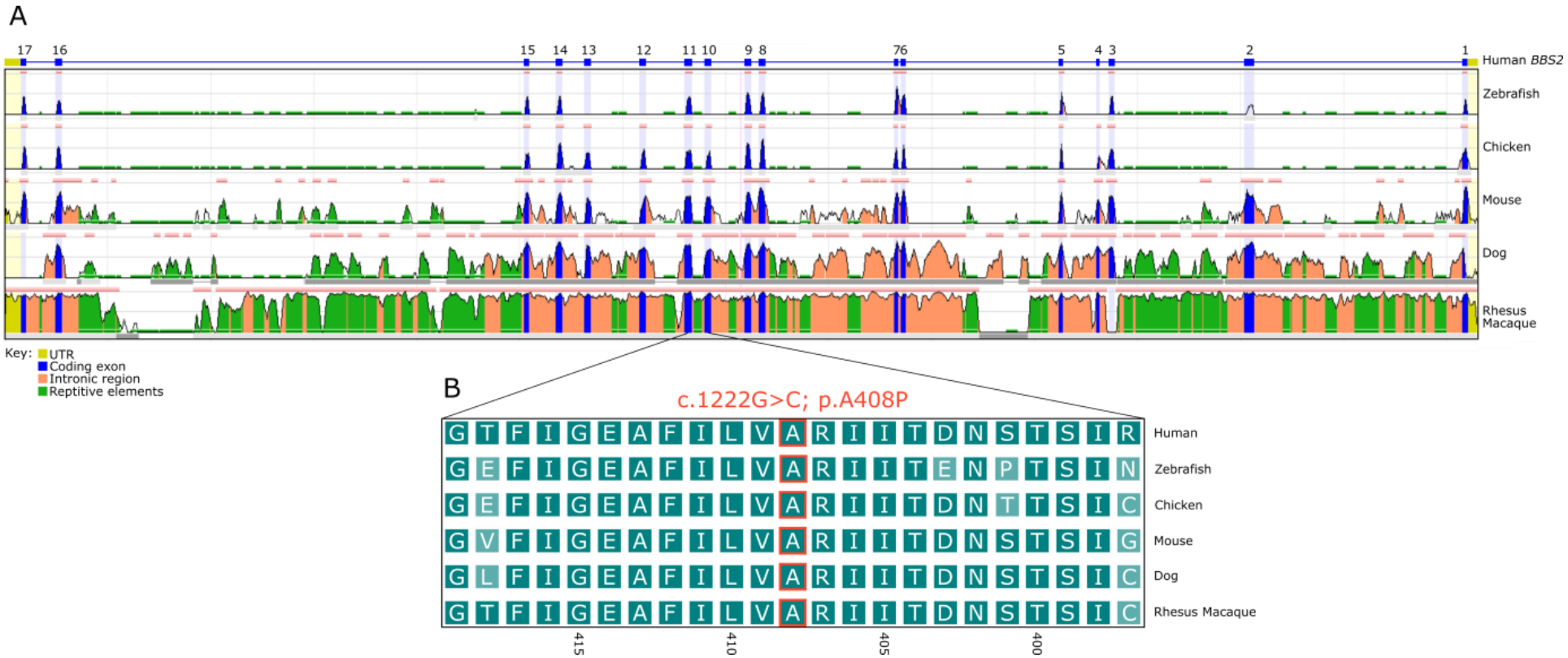


Figure 5.5: Display of *BBS2* conservation across species. **(A)** Evolutionary conserved regions of human *BBS2* (GRCh37; chr16:56517591-56553340) are most highly conserved across the rhesus macaque and dog genomes (CanFam2 assembly) followed by the mouse. The chicken and zebrafish constructs show poorer conservation, however the majority of exonic regions are conserved. **(B)** Evolutionary conservation of the amino acid affected by the canine *BBS2* missense variant (p.A408P; alanine boxed in red) and surrounding amino acids across the human, zebrafish, chicken, mouse, dog and rhesus macaque show a high degree of conservation. Amino acids numbered are relative to the dog sequence to identify the amino acid affected by the c.1222G>C; p.A408P SNV.

5.3.8 *BBS2* domains and features

The canine *BBS2* gene encodes a 715 amino acid protein with three domains (Figure 5.6). The N-terminal is predicted to be positioned from amino acid residues 22-120; a mid-terminal from 159-266 and a C-terminal from 270-709 (Pfam; El-Gebali et al, 2019) with a putative coiled-coil domain between 317–344 amino acid residues (UniProt). The c.1222G>C; p.A408P *BBS2* SNV identified in the SS PRA case resides within the predicted C-terminal domain.

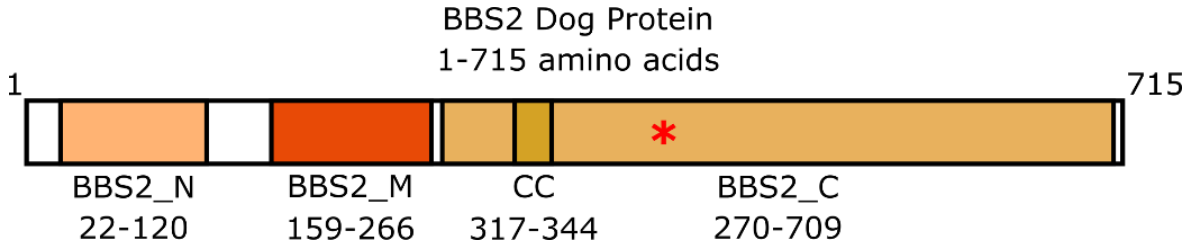


Figure 5.6: BBS2 dog protein. The N-terminal (BBS2_N) is positioned between amino acid residues 22-120. The middle region domain (BBS2_M) is located between amino acids 159-266. The C-terminal (BBS2_C) resides between amino acids 270-709 with a putative coiled-coil domain (CC) from 317-344 amino acids. The BBS2 candidate SNV affects amino acid position 408, as highlighted by a red asterisk.

5.3.9 *BBS2* variant screening

The *BBS2* c.1222G>C SNV was genotyped across 1,386 canids of 155 dog breeds (Appendix 11), 15 cross breeds and eight wolves (Table 5.2), including 505 SS. Of the 505 SS, 91 were submitted to the AHT, of which 78 were homozygous for the wild type allele, 11 were heterozygous and five were homozygous for the *BBS2* SNV. Three-hundred-and-ninety-four SS were submitted to The University of Helsinki, Finland, including one PRA-affected dog, two dogs with suspected PRA and two other retinopathy-affected dogs. These comprised six SS heterozygous for the *BBS2* c.1222G>C SNV and 388 homozygous for the wild type allele. Eight SS were submitted to The Norwegian University of Life Sciences, Norway, of which two were homozygous for the *BBS2* SNV and six were homozygous for the wild type allele. Nine of the 505 SS were from The University of Pennsylvania, USA (including one PRA-affected SS and eight SS related to this case); all these USA samples were homozygous for the wild type allele. Three SS belonged to the DBVDC and were homozygous wild type for the *BBS2* SNV. Taken together, 14 SS were confirmed PRA cases, of which four were homozygous for the *BBS2* SNV and ten were homozygous for the wild type allele. The four cases homozygous for the SNV included two SS in the original litter, and two SS from the Norwegian cohort. Pedigree information from the two dogs from the Norwegian cohort showed they were non-related within five generations to the WGS case, and no photographs or information were available regarding any additional phenotypic information. Genotypes of SS dogs tested for the *BBS2* SNV, that were either confirmed PRA cases or PRA controls known to be clear of PRA or where a PRA diagnosis was unknown, are summarised in Table 5.3.

In addition, the 858 dogs of 154 other breeds, including 136 Border Collie dogs (six PRA-affected), 15 cross breeds and eight wolves that were also genotyped for the *BBS2* SNV were homozygous for the wild type allele.

Table 5.2: Allele frequency of *BBS2* SNV in multiple breeds of dog.

Breed	*BBS2* +/+	*BBS2* +/−	*BBS2* −/−	Allele Frequency
Shetland Sheepdog (SS)	481	17	7	0.031
SS *	481	13	4	0.021
Cross breeds	15	0	0	0
Wolves	8	0	0	0
154 other breeds	858	0	0	0

* excluding third generation relatives to probands from known pedigree information.

Table 5.3: Genotype distribution of the *BBS2* SNV in SS confirmed PRA cases and SS PRA controls, or SS with an unknown PRA diagnosis.

Variant	SS PRA cases			SS PRA controls or unknown PRA diagnoses		
	Alt/Alt	Alt/Ref	Ref/Ref	Alt/Alt	Alt/Ref	Ref/Ref
BBS2 c.1222G>C SNV	4	0	10	3	17	471

Genotyping of two SS, aged one year (Figure 5.7) and 7.6 years old (Figure 5.8), that exhibited the same distinctive characteristics as the initial case (upturned nose, unusual coat and dental defects) revealed that both are homozygous for the SNV. The 7.6-year-old *BBS2* homozygote underwent ophthalmoscopic examination by a board-certified ophthalmologist (who was also a BVA/KC/ISDS panellist), however no clinical signs of PRA were observed. The one-year-old SS did not undergo ophthalmic examination. Pedigree information for these two additional *BBS2* homozygotes shows that both dogs share common ancestry to the initial SS litter of three BBS2 homozygotes (Figure 5.9). The pedigree also highlights third generation relatives to the proband, which were excluded in determining allele frequencies to be more representative of the general SS population.

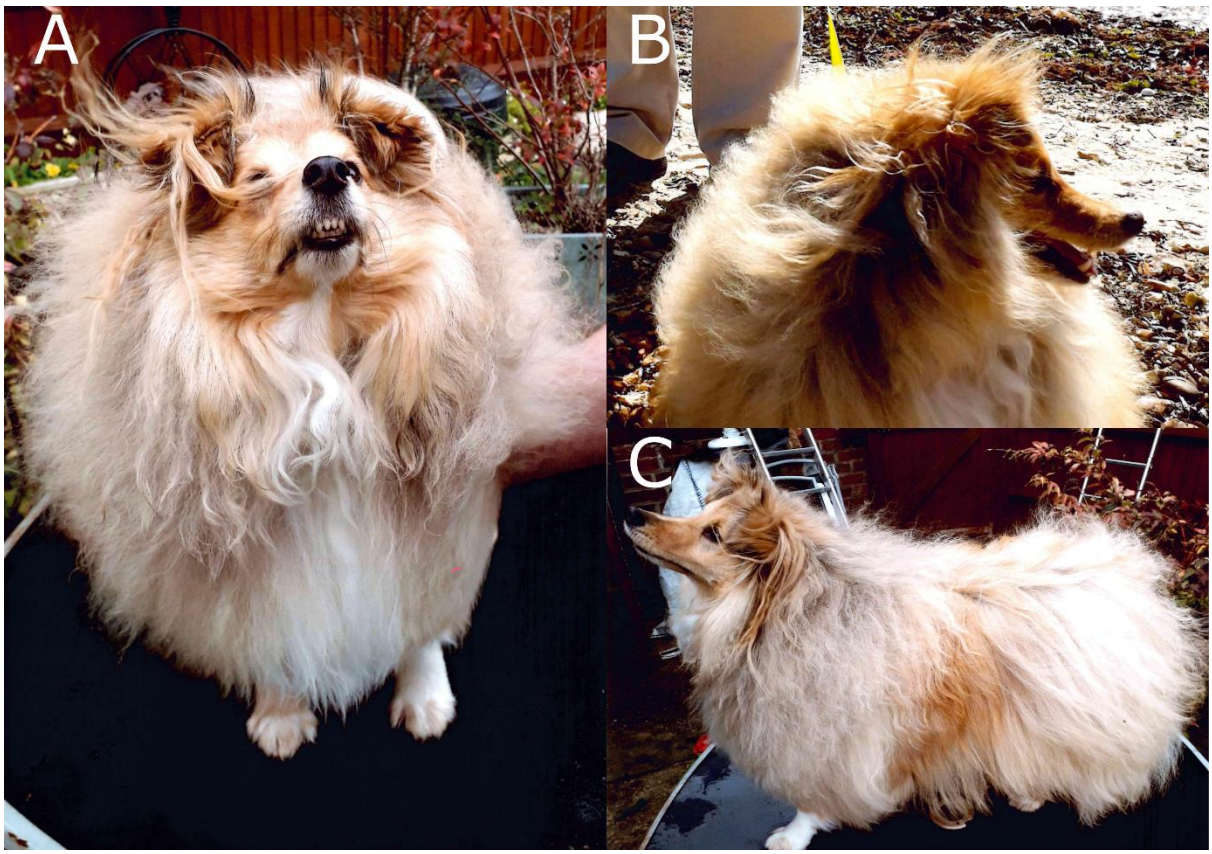


Figure 5.7: Photographs of the one-year-old SS homozygous for the *BBS2* variant displays distinctive features akin to the initial female SS PRA case in the study **(A-C)**.

Figure 5.8: Photographs of the 7.6-year-old *BBS2* homozygote SS showing a wider snout **(A)**, slightly wavy coat texture and a slight upturned nose **(B, C)**.

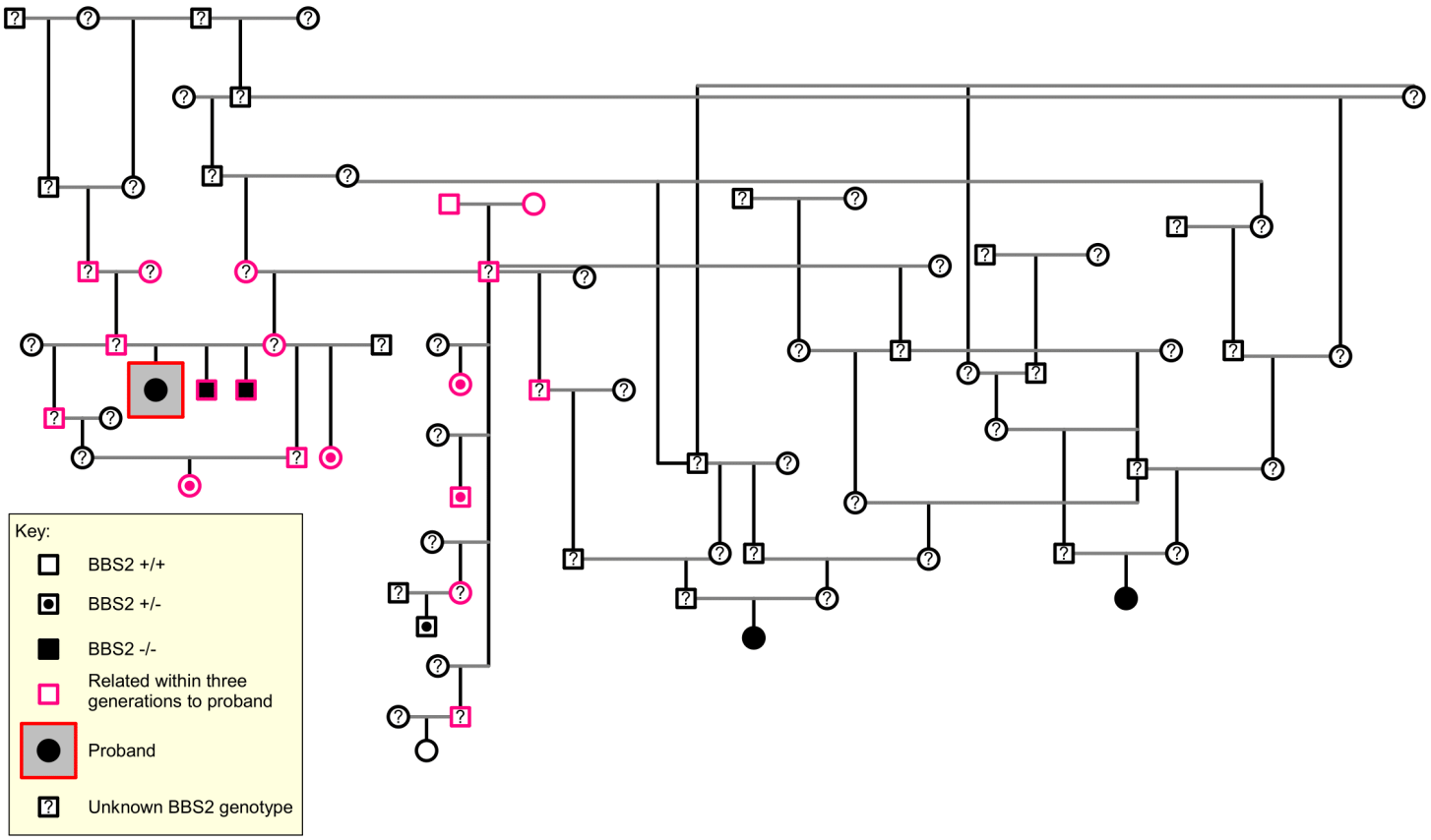


Figure 5.9: Pedigree drawing of *BBS2* homozygous SS shows that there is shared ancestry between five of the *BBS2* homozygotes. Male dogs are shown as a square symbol and female dogs as a circle symbol. Individuals coloured pink are related within three generations to the proband sent for WGS.

Of the seven dogs homozygous for the *BBS2* c.1222G>C SNV, photographs that were obtained for three dogs show a striking upturned nose (Figure 5.10A, B, C) with two displaying a slightly more pronounced nose (Figure 5.10A, C). Comparison of these dogs against a SS that is a half-sibling to the initial PRA case; heterozygous for the SNV; and exhibits a typical Sheltie nose shape and coat (Figure 5.10D) suggests heterozygotes do not display this disease phenotype.

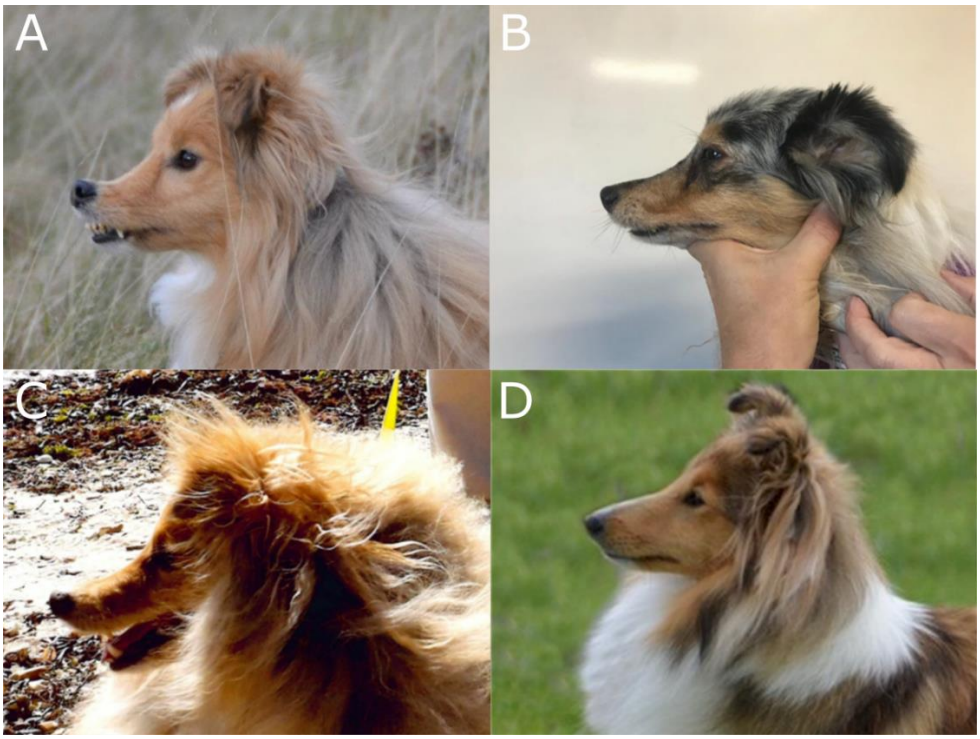


Figure 5.10: Profile view photographs of SS dogs tested for the *BBS2* SNV. SS in photographs **(A)**, **(B)** and **(C)** are homozygous for the *BBS2* c.1222G>C SNV and exhibit an upturned nose. Dog **(D)** is heterozygous for the SNV and was clinically clear of PRA at the age of 9.2 years.

5.3.10 Additional BBS-gene variants

The VCF output file for the SS generated 510,752 variants of all effect scores (1-5), of which 174 variants reside in one of the 21 genes implicated in human BBS (including the *BBS2* candidate variant). Of the 174 variants, three variants are situated in the 3'-UTR region (of *BBS10*), 17 are intergenic variants and 152 intronic variants. The only exonic variants are the c.1222G>C SNV in *BBS2*, present only in the SS, and a missense variant in *CEP290* (c.3011C>T; p.S1003L). In 199 genomes in the AHT genome bank, the *CEP290* variant was homozygous in the SS as well as three other breeds (one Border Collie crossbreed, one Bulldog and one Smooth Collie) and heterozygous in four additional breeds (one Bulldog, one Bearded Collie, one Rough Collie and a Swedish Vallhund dog). In the DBVDC cohort, the *CEP290* variant was homozygous in dogs of the following breeds: Cardigan Welsh Corgi (*n* = 1), Collie (*n* = 1), Scottish Deerhound (*n* = 2) and SS (*n* = 2), and heterozygous in: American Bulldog (*n* = 1), French Bulldog (*n* = 1), German Shepherd cross (*n* = 1), Greyhound (*n* = 1), Irish Wolfhound (*n* = 1), Malinois (*n* = 2), Newfoundland (*n* = 1), Old English

Sheepdog (*n* = 1), Scottish Deerhound (*n* = 1), Pembroke Welsh Corgi (*n* = 1), SS (*n* = 1) and St Bernard (*n* = 1). The *CEP290* SNV is present in dbSNP (rs851859358) with a minor allele frequency of 0.021, but no phenotype information is available. The variant is tolerated by SIFT with a score of 0.26. Five SS homozygous for the *BBS2* SNV, one SS heterozygous for the SNV and one SS homozygous for the wild type/ common allele were genotyped for the *CEP290* c.3011C>T SNV. All eight SS were homozygous for the alternate 'T' allele at this position.

5.4 Discussion

Using WGS of a single SS PRA-affected dog compared to 176 non-breed matched controls, a SNV in the *BBS2* gene located on CANFA2 has been identified. *BBS2* is a strong candidate gene for canine PRA as it has been implicated in the human BBS, of which retinal degeneration is a primary clinical sign. The candidate variant was absent from 858 dogs of 154 other breeds, 15 cross breeds and eight wolves, indicating that the variant was probably unique to the SS breed. Further to the PRA case sent for WGS, six additional SS were homozygous for the SNV, including two full siblings to the WGS case. In these additional homozygotes, where available, clinical information and photographs confirmed a PRA diagnosis and/or features of an upturned nose, abnormally textured coat and dental abnormalities. The involvement of *BBS2* in human BBS patients sharing similar clinical features presents *BBS2* as an interesting gene to consider a syndromic form of PRA in the SS.

Human BBS is classically an autosomal recessive multisystem ciliopathy. There are a wide array of clinical manifestations, whereby a BBS diagnosis is based on the presence of at least four primary features, or three primary and two (or more) secondary features. In human patients, primary features of the condition include retinal degeneration; polydactyly; obesity; genital abnormalities; renal abnormalities; and learning difficulties. A diverse spectrum of secondary features manifest, including speech and developmental delays, diabetes; dental defects; congenital heart disease; brachydactyly; ataxia; and craniofacial abnormalities, described below (Forsythe et al, 2018). The most common feature that typically defines BBS is RP, observed in 93% of patients (Beales et al, 1999; Forsythe & Beales, 2013) where a cone-rod or rod-cone degeneration can be present. The age of diagnosis is variable and, due to a reduction in rod photoreceptor response, night blindness is an initial sign typically apparent between the ages of 5-10 years, although this can present in infancy, with total blindness occurring before the age of 20 years (Beales et al, 1999; Forsythe et al, 2018). Canine PRA is clinically akin to human retinal degeneration, specifically RP. Two out of three SS in the initial litter were diagnosed with bilateral retinal degeneration, based on a PRA-like appearance, and additional homozygotes were also clinically affected with PRA. Exceptions to this include the third SS in the initial litter, which was reported blind with PRA-like symptoms by the owner but was not confirmed by a veterinary ophthalmologist, and the two SS unrelated to the initial litter, aged one year and 7.6 years. Where a PRA diagnosis was made, the age of onset of PRA varied between the two related SS cases with the initial female SS diagnosed within a year of a previous clear eye examination, aged 8.7 years, and the second dog in the litter at an earlier diagnosis of 5.5 years. Therefore, the *BBS2* c.1222G>C homozygote that was clinically clear of PRA at the age of 7.6 years, and the one-year-old homozygous SS, will develop this form of PRA within their lifetimes. Routine ophthalmoscopic follow-up of *BBS2* homozygotes will determine disease age of diagnosis.

There is evidence that BBS patients also display unique, subtle craniofacial secondary features (Beales et al, 1999; Lorda-Sanchez et al, 2001; Moore et al, 2005; Tobin et al, 2008). These features include undeveloped cheek bones (malar hypoplasia); a depressed nasal bridge; deep set eyes with short or narrow palpebral fissures; and a wide forehead (Forsythe et al, 2018; Lorda-Sanchez et al, 2001; Moore et al, 2005). The lower jaw can be recessed further back than the upper jaw, thus causing an overbite (retrognathia), and the lower lip slightly everted (Forsythe et al, 2018). Knowledge of these facial features may help facilitate an earlier diagnosis of BBS before other primary and secondary features present. Dentofacial defects are also observed in > 50% of affected individuals and include dental crowding (~ 46% of cases), a high arched palate (~ 35% of cases) and short dental roots (Borgström et al, 1996; Ferreira do Amaral et al, 2014; Panny et al, 2017). Photographs of three SS homozygous for the *BBS2* SNV in the present study (Figure 5.10A-C) show a distinctive upturned nose shape that is uncharacteristic of the breed, possibly due to a depressed nasal bridge as described in human BBS, and some dental abnormalities, potentially with dental crowding.

Another commonly observed primary feature of BBS is obesity, occurring in approximately 72-92% of patients (Forsythe & Beales, 2013). Obesity is a result of an imbalance in energy ingested and energy expended, where calorie intake exceeds energy expenditure. Energy balance is regulated by hormonal and neural signals in the central nervous system through hormones such as insulin and leptin acting as adiposity signals, involved in regulating energy intake and expenditure (Morton et al, 2006). Leptin originates from the adipose tissue and is released into the blood in proportion to body fat mass, i.e. if more adipose tissue is present, more leptin is released. Leptin is involved in the neuroendocrine circuit to control appetite and homeostasis, interacting with neurons in the brain that decrease food intake such as orexigenic agouti-related peptide (AgRP) and proopiomelanocortin (POMC) derived neurons (Cancello et al, 2004; Rahmouni et al, 2008). The regulation of POMC and AgRP neurons relies on a functional BBSome to maintain energy homeostasis (Guo et al, 2019). Interestingly, mutations in the *POMC* gene are associated with obesity, specifically with increased food motivation, in humans (Challis et al, 2002; Kühnen et al, 2016; Lee et al, 2006) as well as Labrador Retriever and Flat-Coated Retriever dogs (Raffan et al, 2016). There is evidence that BBS-affected individuals have an increased visceral adiposity and higher levels of leptin, suggestive of greater leptin resistance and hyperleptinemia (Feuillan et al, 2011). Similar findings of leptin resistance associated with obesity have also been displayed in *BBS1*, *BBS2*, *BBS4* and *BBS6* mouse models, where BBS-null mice had an increased appetite for food consumption, reduced locomotor activity and showed increased levels of leptin (Davis et al, 2007; Rahmouni et al, 2008). This led to speculation that leptin resistance in human BBS patients may be facilitated by defects in the BBSome that contribute to the regulation of leptin receptors and downstream targets (Feuillan et al, 2011). BBS-knockout mouse studies established that BBS proteins interact with leptin receptors and play a role in

trafficking, thus aberrant leptin receptor trafficking and reduced signalling leads to leptin resistance and is a major cause of obesity in BBS (Seo et al, 2009). Furthermore, dysfunction of leptin receptor signalling is associated with reduced expression of *POMC*. POMC neurons in the hypothalamus of *Bbs2*$^{-/-}$, *Bbs4*$^{-/-}$ and *Bbs6*$^{-/-}$ mice do not respond to changes in peripheral energy balance, indicating that *POMC* is regulated by leptin and critical for the control of energy balance. Thus, impairment of POMC neurons downstream of leptin receptor signalling is probably the main cause of obesity in BBS mice (Seo et al, 2009). Clinical history of two of the three SS in the original litter, all of which are homozygous for the *BBS2* SNV, revealed that both dogs were on a low-fat diet for much of their adult life, with one classed as severely obese. A *BBS2* homozygote identified through additional genotyping had not started showing signs of PRA at the age of 7.6 years, however it had showed signs of an increased motivation for food, described by the owner as a having a "voracious" appetite. It is plausible that these dogs have dysregulation of satiety hormones such as leptin due to dysfunction of BBS2 in the BBSome, resulting in an increased food intake and consequently leading to weight gain or obesity. Due to limited clinical information of other SS that are homozygous for the *BBS2* SNV, sufficient evidence has yet been found to support the hypothesis of an association between the *BBS2* SNV and an increased appetite or obesity. To determine whether weight is a contributing factor of the *BBS2* canine phenotype, or whether these homozygotes were overweight due to overfeeding by owners, requires further study of *BBS2* homozygotes.

Renal involvement is highly variable and is a major contributor to morbidity and mortality in BBS patients, with an estimated prevalence of 53-82% (Forsythe et al, 2017; Imhoff et al, 2011). Defects include cystic tubular disease, anatomical deformities and renal dysplasia, leading to kidney failure. Early signs of kidney dysfunction are polyuria or polydipsia with urinary concentration defects and a large proportion of cases exhibit renal lesions (Tieder et al, 1982). Variability in renal involvement is highlighted by research exploring the prevalence of chronic kidney disease in BBS patients, which indicated those with mutations in *BBS2*, *BBS10* or *BBS12* were more likely to exhibit severe kidney disease than patients with mutations in the *BBS1* gene (Forsythe et al, 2017). The involvement of renal anomalies in BBS may be due to several processes in the renal system, whereby cilia are present and functionally significant. Primary cilia are present on the surface of epithelial cells of the nephron and collecting duct in the renal cortex of the kidney and play a role in epithelial homeostasis and repair (reviewed in Deane et al, 2013). Information on kidney health was not available for all SS dogs in the present study. However, of the three SS littermates in the initial litter, two displayed kidney malfunctions in their later years of life, most of which were referred to in veterinary records as kidney failure. For one case (female SS), renal changes were exhibited at the age of 13.3 years, and later, possible chronic renal degeneration was remarked. A second littermate (male A) consistently showed increased levels of urea and creatine in the urine with concerns about renal function and chronic renal failure,

despite being on a kidney support diet and medication for renal disease.. By the age of 9.3 years, kidney disease had progressed significantly and shortly after the owner elected for euthanasia.

While human BBS phenotypes are characterised by defects in 21 genes, all involved in the functioning of primary cilia (Table 5.4), many encoding genes involved in BBS specifically regulate the machinery of the multi-subunit complex termed the BBSome. The BBSome consists of eight highly conserved proteins: BBS1, BBS2, BBS4, BBS5, BBS7, BBS8, BBS9, and BBS18/BBIP10, which function in primary cilium biogenesis (Loktev et al, 2008; Nachury et al, 2007) (Figure 5.11). Additional components include the small Arf-like GTPase comprising BBS3/Arl6, which recruits the BBSome onto membranes for coat assembly (Jin et al, 2010); LZTFL1/BBS17, which is involved in regulating entry of the BBSome into cilia (Seo et al, 2011); and a BBS/CCT (chaperonin containing T-complex protein-1) complex comprising BBS6, BBS10 and BBS12, required for BBSome assembly (Seo et al, 2010) (Figure 5.11). Other known BBS genes: BBS11, BBS13, BBS15, BBS16 and BBS19-21 (Aldahmesh et al, 2014; Chiang et al, 2006; Heon et al, 2016; Kim et al, 2010; Leitch et al, 2008; Lindstrand et al, 2016; Otto et al, 2010; Xing et al, 2014) are not as closely associated with the BBSome, and their role in cilia is not completely characterised.

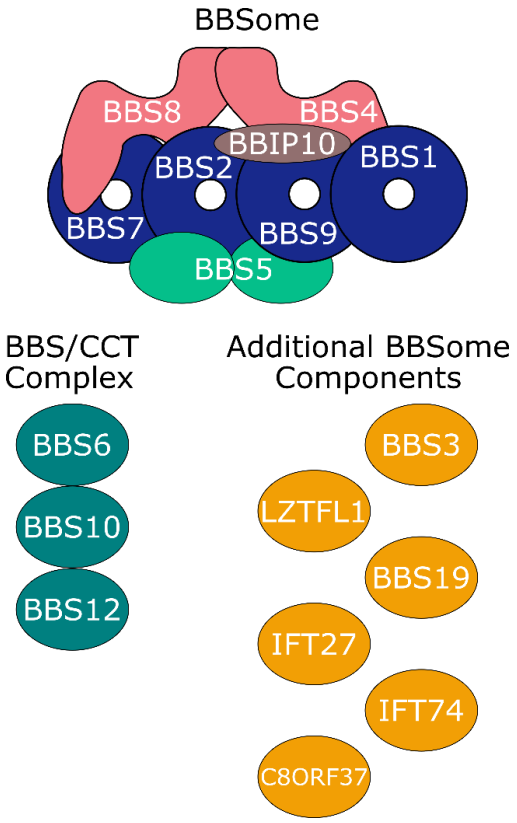


Figure 5.11: Schematic diagram of the BBSome and interacting components. The BBS/CCT complex and additional components shown interact with the BBSome (adapted from Jin & Nachury, 2009).

BBS2, BBS7 and BBS9 form the core of the BBSome, termed the BBSome core complex, important in BBSome maturation (Zhang et al, 2012). Structurally, BBS2, BBS7 and BBS9 are homologous proteins that share distinct protein folding and domain organisation: a β-propeller fold within the N-terminal and a coiled-coil and a γ-adaptin ear (GAE) motif accompanied by an α/ β platform domain in their C-terminal domain (Jin et al, 2010). Within the BBS2-7-9 subcomplex, BBS2 is located between BBS7 and BBS9 with the α-helical domain of BBS2 in contact with that of BBS9 (Ludlam et al, 2019). Thus, BBS2 strongly interacts with BBS9 in this complex.

Table 5.4: Genes associated with BBS in humans. An asterisk (*) indicates those genes that have also been implicated in canine PRAs.

BBS type	Gene	Study reference
BBS1	*BBS1*	(Mykytyn et al, 2002)
BBS2	*BBS2*	(Nishimura et al, 2001)
BBS3	*ARL6*	(Chiang et al, 2004)
BBS4	*BBS4**	(Mykytyn et al, 2001)
BBS5	*BBS5*	(Li et al, 2004)
BBS6	*MKKS*	(Katsanis et al, 2000; Slavotinek et al, 2000)
BBS7	*BBS7*	(Badano et al, 2003)
BBS8	*TTC8**	(Ansley et al, 2003)
BBS9	*BBS9*	(Nishimura et al, 2005)
BBS10	*BBS10*	(Stoetzel et al, 2006)
BBS11	*TRIM32*	(Chiang et al, 2006)
BBS12	*BBS12*	(Stoetzel et al, 2007)
BBS13	*MKS1*	(Leitch et al, 2008)
BBS14	*CEP290*	(Leitch et al, 2008)
BBS15	*WDPCP*	(Kim et al, 2010)
BBS16	*SDCCA8*	(Otto et al, 2010)
BBS17	*LZTFL1*	(Marion et al, 2012)
BBS18	*BBIP10*	(Scheidecker et al, 2014)
BBS19	*IFT27*	(Aldahmesh et al, 2014)
BBS20	*IFT172*	(Lindstrand et al, 2016)
BBS21	*C8orf37*	(Heon et al, 2016)

BBS2-associated variants have been identified in patients with BBS type-2 (BBS2) (Beales et al, 1997; Bee et al, 2015; Castro-Sanchez et al, 2017; Chen et al, 2011; Ece Solmaz et al, 2015; Katsanis, 2001; Nishimura et al, 2001) as well as non-syndromic RP cases (Shevach et al, 2015). The *BBS2* locus was initially mapped in humans to chromosome 16q (Kwitek-Black et al, 1993) and led to the identification of a homozygous 1-bp deletion in exon 8 of *BBS2* in a family with BBS, and a non-synonymous SNV in a second BBS proband (Nishimura et al, 2001). Further to these, several novel homozygous and compound heterozygous nonsense, frameshift and missense mutations in *BBS2* were identified (Katsanis, 2001). Another study utilised WES analysis to identify homozygous variants in *BBS2* as a genetic diagnosis for two BBS patients presenting with clinical features including RP, obesity, polydactyly and craniofacial anomalies (Castro-Sanchez et al, 2017). Nonsense variants were revealed in exons 5 and 16, respectively, and were deemed to be the cause of their ciliopathy-like phenotypes. An earlier *BBS2* mutation identified by Katsanis et al. (2001), causing an amino acid substitution of arginine to proline at residue 632 (R632P), was also identified in BBS patients in later studies (Bin et al, 2009; Chen et al, 2011). Binding experiments to investigate the impact of the R632P substitution revealed that the mutation is located in the α-helical domain of BBS2 (Ludlam et al, 2019). Ludlam et al. (2019) demonstrated that BBS2 interacts with BBS7, forming a tight dimer that associates with the rest of the BBSome hexameric subcomplex (comprised of BBS1, 4, 5, 8, 9, 18). Therefore, the R632P substitution disrupts BBS2-BBS9 interaction and, in turn, BBSome assembly, preventing the BBS2-BBS7 dimer from associating with the BBSome hexameric subcomplex (Ludlam et al, 2019). Given that the *BBS2* c.1222G>C; p.A408P SNV in the present study is predicted to be positioned within the C-terminal of BBS2, and considering the results of aforementioned studies establishing the presence of the α-helical domain of BBS2 within the C-terminal, it may be hypothesised that this candidate SNV in the SS is detrimental to the role of BBS2 in the canine BBS2-7-9 complex; comparable to the human R632P substitution, i.e. that BBSome assembly is disrupted by the *BBS2* c.1222G>C; p.A408P SNV, which may impact ciliary transport machinery in the complex.

Although BBS is generally of autosomal recessive inheritance, there have been suggestions that the disorder can manifest as an oligogenic disease model, where in some instances BBS is not inherited as a simple single-gene Mendelian disorder nor does it fit a complex disease model. Instead, an oligogenic phenotype manifests when mutations are present in a small number of genes that interact with one another. Triallelic inheritance is an example of an oligogenic model, whereby two mutant alleles at one locus and one mutant allele at another locus are required for disease manifestation; exemplified by mutations in *BBS2* and *BBS6*. A study by Katsanis et al. (2001) found that 47% of patients with *BBS2* mutations also harboured a third mutation in a second BBS gene. Thus, genetic modifiers may impact homozygous BBS gene mutations and lead to variation in BBS phenotypes. To assess the possibility of an oligogenic pattern of inheritance in the current study, further exploration of variants in all 21 known BBS genes was conducted in the

affected SS WGS. No variants were identified in the *BBS6* gene previously demonstrated to be involved in triallelic inheritance with *BBS2* in humans. Despite the identification of many intronic variants across other BBS genes in the PRA-affected SS genome, the only notable exonic variant found to segregate, in addition to the *BBS2* c.1222G>C; p.A408P SNV, was a c.3011C>T; p.S1003L SNV in *CEP290. CEP290* is associated with BBS14 in humans. This missense variant is homozygous in five SS homozygous for the *BBS2* SNV as well as SS heterozygous for the SNV and homozygous for the wild type allele. The *CEP290* SNV is also present (homozygous and heterozygous) in multiple non-breed matched controls. Therefore, while it may not be pathogenic on its own, it could be a rare polymorphism that contributes to the BBS phenotype when inherited in combination with the *BBS2* SNV. This seems unlikely as this SNV is present across multiple dog breeds, and as each breed is an isolated breeding population this is more likely to be an ancestral polymorphism. Current evidence suggests that the BBS-phenotype in the SS is a single-gene recessive disorder caused by the exonic SNV in *BBS2*. To exclude the involvement of intronic variants as potential oligogenic contributors such as triallelic or modifier mutations, targeted sequencing of all BBS genes across numerous SS PRA cases homozygous for the *BBS2* c.1222G>C; SNV and control SS dogs would be required. Furthermore, this study did not assess the functional impact of intronic variants including potential intronic cryptic/pseudo splice site variants.

Studies of *Bbs2*-null mice show that the absence of *BBS2* expression results in retinal degeneration, male infertility, renal cysts and obesity. Consistent with the Sheltie *BBS2* homozygotes, normal retina development and function is present in young *Bbs2*-null mice followed by progressive loss of photoreceptors: the primary ciliated cells of the retina (Nishimura et al, 2004). Flagella formation during spermatogenesis in *Bbs2*-null male mice was impaired, a feature that was also demonstrated in human BBS-2 patients and *Bbs*-4 null mice (Mykytyn et al, 2004). These findings highlight the involvement of motile cilia in addition to primary cilia in BBS phenotypes. Nishimura et al. (2004) demonstrated that although *Bbs2*-null mice can assemble primary cilia of renal tubule cells, these cilia are tapered and, as with polycystic kidney disease, renal cyst development is likely to be due to signalling defects in primary cilia. Obesity was observed in *Bbs2*-null mice as a result of an increase in food intake due to a defect in satiety regulation, which may be due to a greater leptin resistance, as previously discussed in human BBS. Further neurological findings included olfactory defects and social dominance deficits, where *Bbs2*-null mice lack social dominance when compared to littermate controls; a similar phenotype that was present in other BBS mouse models (Davis et al, 2007; Eichers et al, 2006). These defects in social dominance were replicated in a dark environment to show that this feature was not exclusively due to reduced vision in *Bbs2*-null mice. These findings suggest that, while BBS2 (and the BBSome) is not crucial for initial cilia assembly, it is indirectly involved in cilia maintenance, ciliogenesis and IFT, whereby the absence of *BBS2* in knockout mice impacts intracellular transport and IFT requirements, which are crucial for the maintenance of photoreceptor outer

segments (Nishimura et al, 2004). Since primary and motile cilia lack the machinery to synthesise proteins, IFT is the bidirectional transport of multi-subunit protein complexes, such as the BBSome, and is crucial in cilia and flagella formation and function. Non-membrane-bound particles are transported in an anterograde (forward) direction along the axonemal doublet microtubules beneath the cilium or flagellar membrane, carrying materials essential for the assembly and maintenance of the flagellar axoneme and membrane. Once these reach the tip of the cilium/flagellum, a recycling step is undertaken in a retrograde (backward) direction where IFT particle movement reverses to return to the base of the cilium/flagellum amongst IFT components in the region surrounding the basal body (Rosenbaum & Witman, 2002). As this IFT mechanism is crucial for cilia and flagella formation and maintenance, defects in IFT are detrimental to primary cilia, such as those in photoreceptor outer segments and renal tubule cells, and motile flagella. The multisystem phenotype demonstrated in SS *BBS2* homozygotes may be explained by similar mechanisms described in BBS mouse models.

Previous research demonstrates similarities between BBS-null mice and canine PRA phenotypes, specifically in the Hungarian Puli dog where a nonsense SNV in *BBS4* causes a novel syndromic PRA with phenotypes comparable to *Bbs4*-null mice (Chew et al, 2017). In addition to PRA, *BBS4*-affected Hungarian Puli dogs displayed spermatozoa flagella defects (albeit less severely affected than of that observed in *Bbs4*-null mice) as well as obesity. This was the first implied BBS-like phenotype reported in dogs. Prior to this, the only BBS gene implicated in canine PRA was the *TTC8* gene, also known as *BBS8*, in Golden Retriever dogs (Downs et al, 2014), which has also been implicated in human BBS (Ansley et al, 2003) and non-syndromic RP in humans (Goyal et al, 2016; Riazuddin et al, 2010). However, dogs in the *TTC8* study were not reported to have any abnormal clinical features apart from PRA. There has since been evidence to suggest that the Golden Retriever *TTC8* variant causes a BBS phenotype in homozygous dogs. A Labrador Retriever that is homozygous for the Golden Retriever *TTC8* PRA-associated variant displayed clinical signs of PRA accompanied with an upturned nose, abnormal coat and obesity; comparable to the *BBS2*-homozyous SS in this study (personal communication, Pearce-Kelling, unpublished work). No information regarding spermatozoa flagella potency in SS male dogs was available to conclude if fertility defects are present. However, for one of the *BBS2* homozygous female dogs, the owner did report that the bitch did not have a season until much later in life, around the age of five years. This provides further exploration in future studies of canine *BBS2* models.

The Kennel Club, UK outlines a breed standard for the Sheltie to maintain certain characteristics and traits to ensure new generations are fit for function. The breed standard states: the head should be refined with no exaggerated features with a long, blunt wedge tapering from ear to nose when viewed from a top or side view. Cheeks should be flat, merging smoothly into a well-rounded muzzle and the skull and muzzle to be of equal length. Jaws should be level, clean and strong with

a well-developed underjaw and lips tight. Upper teeth should closely overlap the lower teeth, setting square to the jaws with preferably a full set of 42 properly placed teeth. Breed standard recognised coat colours for the SS are sable, tricolour, blue merle, black and white and black and tan. The coat should be comprised of a double coat with a soft short undercoat and an outer coat of long, harsh-textured and straight hair. The hair around the face should be smooth and the coat should fit the body but not detract from the body outline (The Kennel Club UK, 2019b). Findings from this present study demonstrate that physical features of SS *BBS2* homozygotes do not fit within these criteria. It is plausible that the fullness of the coat may make it difficult to see if a SS is over- or underweight and to track weight appropriately. There is limited evidence showing the prevalence of obesity in SS: one study in the 1980s showed that the UK Shelties were one amongst six other breeds to have a pre-disposition to obesity (Edney & Smith, 1986). A later study in 2006 on the US population indicated that 57% of SS were overweight and an obesity prevalence of 10% (Lund et al, 2006). It is not well documented how this currently stands in the UK population of SS. As previously discussed, obesity is a secondary feature of BBS, leading to the hypothesis that the *BBS2* SNV contributes to the regulation of energy expenditure and satiety hormones in *BBS2* homozygotes. This proposed syndromic PRA in the SS presents with distinct physical features. In theory, if any breeders regarded this physical look as appealing, deciding to selectively breed for this nose shape or coat texture, a PRA-associated phenotype could be inadvertently selected for, posing a detrimental effect on breed health. History shows how actively breeding for certain characteristics and physical traits has altered the appearance of certain dog breeds, and this is apparent when comparing images of certain breeds over the past century. Examples include the Bull Terrier, Boxer and Dachshund, where certain physical characteristics have become more exaggerated over time, including nose shape, skull morphology and body length. Breeding choices based on physical appearance can have implications not only on breeding standards, but breed health, as illustrated with brachycephalic breeds. Another example of breeding for a specific phenotype that inadvertently has led to a recessive, detrimental mutation is demonstrated in the Petit Basset Griffon Vendéen and Shar-Pei dog breeds. Mutations in the *ADAMTS17* gene are associated with primary open angle glaucoma (POAG), a chronic eye condition characterised by a small, sustained rise in intraocular pressure and lens subluxation, in the Petit Basset Griffon Vendéen (Forman et al, 2015); and POAG with primary lens luxation in the Shar-Pei (Oliver et al, 2018). Both breeds are short in stature and an association between height and POAG genotype has been identified (Jeanes et al, 2019). Thus, by selecting for dogs that are shorter in height to adhere to specific breed standards, breeders have inadvertently led to an increase in *ADAMTS17* mutation frequencies, thereby contributing to an increased incidence of POAG in these breeds.

As with the *BBS4*-homozygous Hungarian Puli dogs, it is plausible that the *BBS2* SNV in the current study causes a syndromic PRA in the SS. Typically, when studying canine PRA, a diagnosis is limited to ophthalmoscopic evaluation. Distinguishing between a canine non-syndromic and

syndromic condition requires recognition of multiple clinical signs, where additional features can be challenging to quantify or assess. Changes in social dominance, olfactory defects and learning or developmental delay are some of the behavioural and neurological features displayed in BBS human patients and mouse models. Further to this, obesity is a common problem in dogs, whereby 39-65% of dogs are classified as overweight and between 9-20% are obese (Courcier et al, 2010; German et al, 2018), therefore its association with a genetic disorder can easily be overlooked. Physical facets including craniofacial or dental features are more obvious; however renal deficits would need to be monitored to characterise. In future studies of canine PRA, the involvement of BBS genes should be considered, especially when only ophthalmoscopic information is gathered for cases and information on physical or additional characteristics is limited.

As previously discussed, although there are PRA-associated mutations that have been detected across multiple diverse breeds e.g. PRCD (Zangerl et al, 2006) and RCD4 (Downs et al, 2013), the majority of canine PRAs previously identified are unique to individual breeds. While candidate variants cannot be excluded based on their presence in additional breeds alone, the identification of a variant across individuals of multiple breeds can aid variant filtration. A variant identified in multiple breeds is more likely to be a shared ancestral polymorphism than a causative variant, especially when the variant is present in breeds for which there is no evidence that the breed is predisposed to PRA. Additional evidence must be gathered to formally include/exclude variants from analyses such as the effect on the protein sequence and whether the gene is known to be expressed in retinal tissue.

In the present study, the c.1222G>C SNV is situated in a compelling gene; segregated appropriately in the SS breed; has a predicted severe impact on the protein; and is not shared amongst multiple breeds, endorsing it as the top candidate variant. Genotyping of dogs across 155 purebreds revealed that the *BBS2* c.1222G>C SNV was only found in SS dogs, suggesting it may be private to the SS breed. The allele frequency, excluding third generation relatives to probands from known pedigree information, of 0.021 indicates this is a rare variant and at a low frequency in the current SS population. Despite this, sample collection shows the variant is still present in the current UK population, thus the use of a DNA test (recently launched by the AHT) will aid in reducing and eliminating this form of PRA from the gene pool, possibly in the next few generations. Use of a DNA test will also provide genetic information on the status of this PRA-associated mutation in individual dogs before clinical signs present and therefore provide a tool for breeding choices before dogs reach breeding age. Finally, the absence of the *BBS2* c.1222G>C SNV in additional PRA cases in the SS screening cohort suggests at least one other form of PRA in the breed remains unresolved. This provides opportunity to further explore PRA in the SS breed.

5.5 Conclusion

This is the third report to-date of a BBS gene associated with canine PRA. This study highlights the power of WGS using a single case to identify a PRA-associated variant. Due to the involvement of the *BBS2* gene in human BBS phenotypes, and knowledge of additional features in *BBS2* homozygotes in the current study, the c.1222G>C; p.A408P SNV is deemed the cause of a syndromic PRA, or a form of canine BBS, in the SS breed. Further follow-up of SS *BBS2* homozygotes is required to understand the spectrum and complexity of additional features, ideally throughout the dog's lifetime with the ability to monitor retinal and renal changes as well as weight tracking and fertility programs. Availability of a DNA test will raise awareness of this form of PRA and encourage owners to utilise the DNA test to reduce allele frequency. Utilisation of such a diagnostic test within the SS breeding community will reduce the possibility of future generations becoming affected with this form of PRA.

Chapter 6 General Discussion

Despite the identification of 27 canine PRA-associated mutations in more than 100 dog breeds, there are still many genetically unresolved cases. This formed the foundation of this thesis, with four main aims to:

1. Identify breeds affected with PRA for which the causal mutation is unknown.
2. Elucidate mutations associated with PRA in each breed under investigation, using whole genome investigations, including GWAS and WGS technologies, identifying the most efficient mutation discovery routes for differently structured proband groups.
3. Develop commercial DNA tests for all mutations identified.
4. Offer novel insights for consideration in human retinal degeneration research.

These aims have been achieved, identifying three dog breeds affected with PRA (GS, LA and SS) and elucidating variants associated with novel forms of PRA in these breeds. DNA tests have been developed for each breed under investigation to reduce the incidence of these forms of PRA in each dog breed.

At present, 44 genes are associated with canine IRDs including achromatopsia, congenital stationary night blindness, canine multifocal retinopathy, neuronal ceroid lipofuscinosis, and canine PRAs. Currently 273 genes have been implicated in various human IRDs, including syndromic/systemic retinal diseases, retinopathies, macular degenerations, LCA, BBS and RP; to name a few (RetNet, 2020). Of these 273 human IRD genes, 35 genes are shared across human and canine IRDs. In canine IRD studies, researchers often look to the human field to aid in mutation discovery. Similarly, canine IRD research has the potential to offer novel genes for consideration in human IRD research, with nine genes currently known to be associated only with canine IRDs. Six of these genes lead to a PRA phenotype: *CCDC66* (Dekomien et al, 2010), *SLC4A3* (Downs et al, 2011), *STK38L* (Goldstein et al, 2010), *MAP9* (Forman et al, 2016), *HIVEP3* (Kaukonen et al, 2020) and, as identified in this thesis, *NECAP1* (Hitti et al, 2019).

6.1 Gene interactions and pathways

Of the three genes found to be associated with canine PRAs described in these studies, *IMPG2* and *BBS2* have previously been implicated in human IRDs, and *NECAP1* is a novel retinal candidate gene. The molecular processes in the retina that enable vision, including the phototransduction cascade and the visual cycle, involve a myriad of different proteins in many of the cells of the retina. Many of these proteins are retina-specific or encode genes that are enriched in the retina, therefore mutations in such genes often result in dysfunction of retinal cells and concurrently promote a retinal phenotype. The phototransduction cascade, the visual cycle and the ciliary

transport system are highly specific and significant to the photoreceptor/RPE complex in the retina. As with the majority of IRDs, each gene that harbours a mutation identified in this thesis encodes a protein that participates in one of these pathways in retinal function, including trafficking in vesicle formation in retinal synapses (*NECAP1*), regulation of photoreceptors and the RPE (*IMPG2*), and structural organisation of photoreceptor outer segments and trafficking of primary cilia (*BBS2*).

NECAP1

The GS PRA study demonstrated that *NECAP1* is a plausible candidate gene for retinal disease. *NECAP1* is expressed in the canine retina and several human tissues including brain and retinal tissue, as indicated by the Human Protein Atlas database (Uhlén et al, 2015). The encoding protein, NECAP1, localises to clathrin-coated vesicles where it binds components of adapter protein complexes, including AP-1 and AP-2, to facilitate clathrin mediated endocytosis (CME). Related pathways of *NECAP1* include CME and vesicular transportation in synapses. The STRING v11.0 web server database demonstrates evidence suggesting a functional link between NECAP1 and interacting proteins (Szklarczyk et al, 2019). As an accessory protein to adaptor protein complexes AP-1 and AP-2, NECAP1 interacts with several proteins related to these adaptor protein complexes (AP1G1, AP2M1, AP2A1, AP2A2; Figure 6.1). Previous studies have illustrated the importance of adaptor complexes in photoreceptor development and survival in vertebrate and invertebrate retinas. In Drosophila (Xiong & Bellen, 2013) and mice (Moaven et al, 2013), rhodopsin/arrestin complexes, where arrestin deactivates rhodopsin in the phototransduction cascade preventing accumulation of rhodopsin in photoreceptor cells, interact with AP-2 to regulate photoreceptor survival. In addition, it has been demonstrated that AP-1 in Drosophila plays a crucial role in eye development, where signalling molecules such as Notch are required for the specification of photoreceptor cells in the developing eye and depletion of AP-1 affects the expression of these signalling molecules (Kametaka et al, 2012). This suggests that AP-1 and AP-2, and possibly associated accessory proteins including NECAP1, are important in the intracellular trafficking of cargo molecules required for eye development and photoreceptor survival, where photoreceptor apoptosis results in retinal degeneration. To date, the only reported human disease associated with *NECAP1* loss of function is EIEE (Alazami et al, 2014; Alsahli et al, 2018). Animal IRD models do not always share the phenotype observed in humans where mutations in the same gene are implicated. For example, *RHO* mutations exert different physiological mechanisms in humans as well as different animal models including mice (Lem et al, 1999; Li et al, 1996), pigs (Ross et al, 2012) and dogs (Kijas et al, 2002). However, previous studies aforementioned involving RP mouse models demonstrate that a) CME is present, albeit playing a minor role, in retinal ribbon synapses in photoreceptor and bipolar cells of the retina; and b) adaptor complexes AP-1 and AP-2, of which *NECAP1* is incorporated as an accessory protein to mediate vesicle transportation in synapses, are crucial for photoreceptor survival.

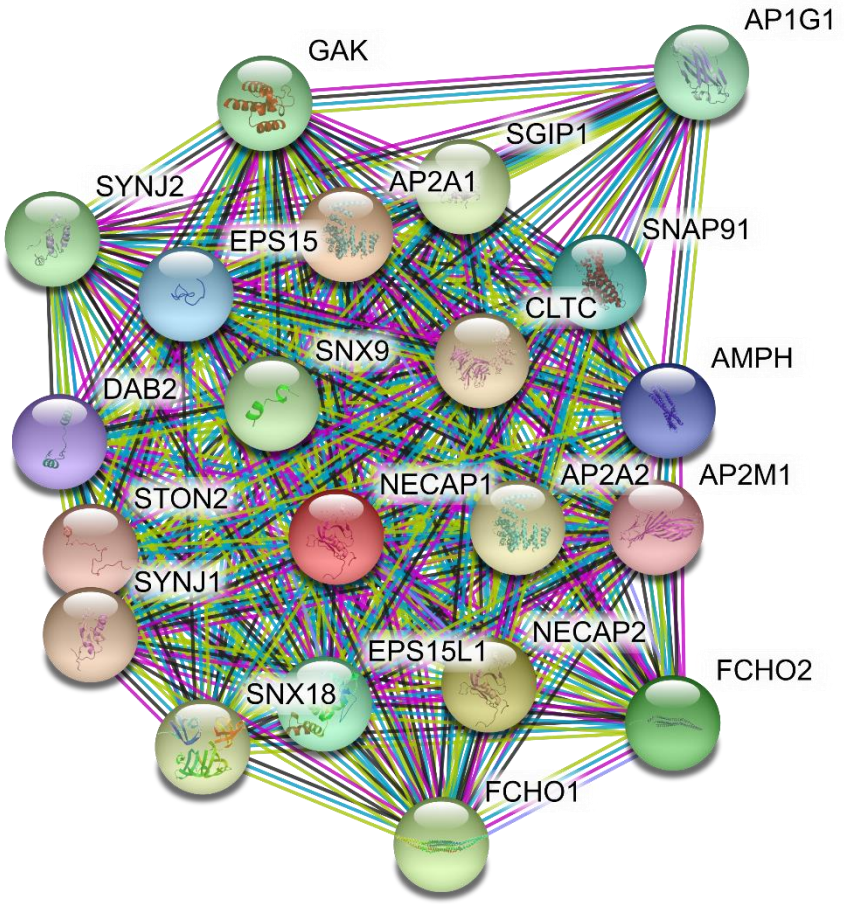


Figure 6.1: Human NECAP1 protein interaction map in STRING v11.0 (Szklarczyk et al, 2019) showing the top 20 interactors with a medium confidence score (0.400). Line colour indicates the types of evidence used in predicting the associations: cooccurrence evidence (blue), experimental evidence (purple), text mining evidence extracted from published literature (yellow), curated database evidence (light blue) and co-expression evidence (black). All underlying evidence can be inspected in dedicated viewers that are accessible from the network.

IMPG2

Previous reports have identified disease-associated variants in *IMPG2* in human patients sharing retinal features with those of canine PRA (Bandah-Rozenfeld et al, 2010; Brandl et al, 2017; Khan & Al Teneiji, 2019), highlighting the gene as a strong candidate for canine retinal disease studies. According to The Human Protein Atlas database, *IMPG2* is one of 310 genes with elevated RNA expression in the human retina and one of 87 genes with at least four-fold higher mRNA level in retina, compared to other tissue types, with known expression in photoreceptor cells (Uhlén et al, 2015). The encoding protein is localised within the interphotoreceptor matrix (IPM) occupying the interface surrounding rod and cone photoreceptors and is a constituent of proteoglycans, which are secreted into the IPM and bind to the RPE. The STRING v11.0 web server database demonstrates evidence suggesting a functional link between the human IMPG2 protein and other proteins involved in the function and maintenance of retinal components including ABCA4,

C2ORF71, FAM161A, PRCD and TTC8 (Figure 6.2; Szklarczyk et al, 2019). Canine PRA-associated mutations have been identified in the genes encoding these proteins: *ABCA4* (Makelainen et al, 2019), *C2ORF71* (Downs et al, 2013), *FAM161A* (Downs & Mellersh, 2014), *PRCD* (Goldstein et al, 2006; Zangerl et al, 2006) and *TTC8* (Downs et al, 2014). All these genes encode proteins that are specific to pathways associated with retinal photoreceptor cells and connecting cilium. Human and canine photoreceptors share similar anatomical structures despite differences in density across areas of the retina, therefore dysfunction of these pathways result in similar retinal phenotypes and concomitantly promote progressive retinal degeneration in both species. In contrast, *NECAP1* interacts with many genes, none of which are known to directly cause a retinal phenotype in humans. This could be because NECAP1 is an accessory protein and interacting genes in the same pathways can compensate for a defective NECAP1 protein.

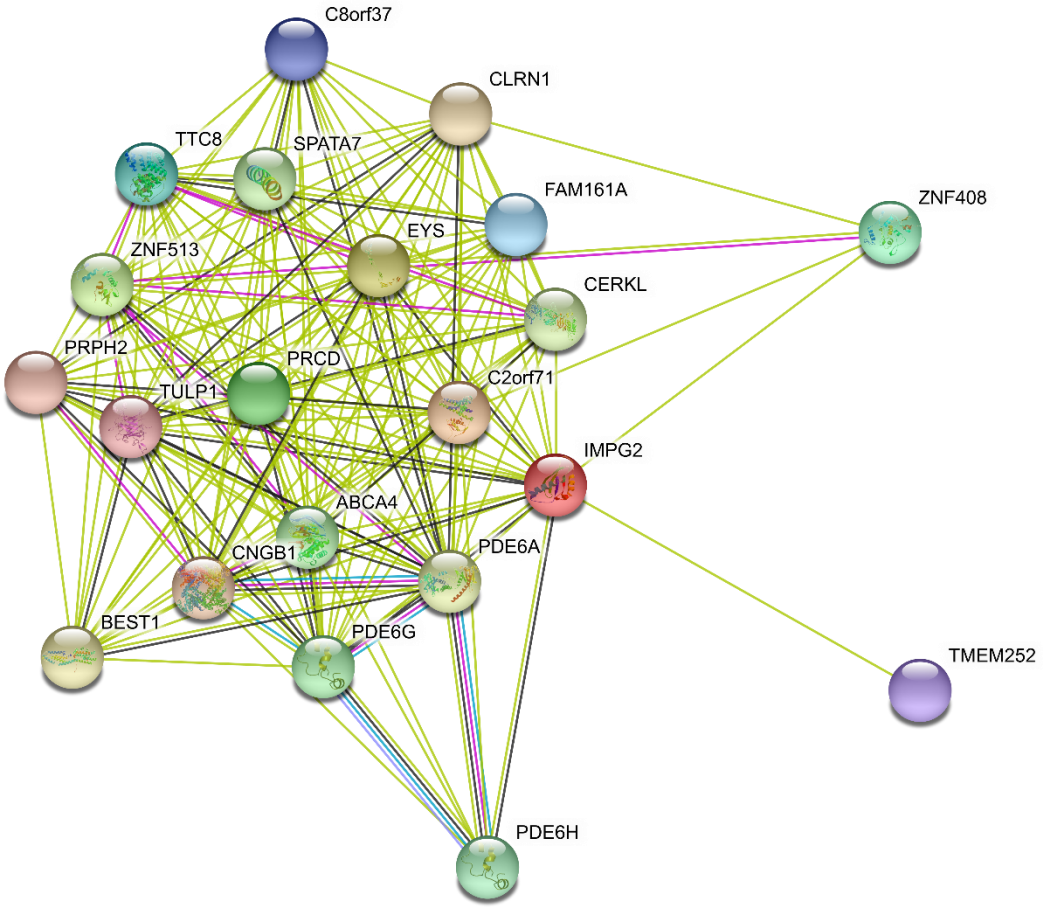


Figure 6.2: Human IMPG2 protein interaction map in STRING v11.0 (Szklarczyk et al, 2019). showing the top 20 interactors with a medium confidence score (0.400). Line colour indicates the types of evidence used in predicting the associations: cooccurrence evidence (blue), experimental evidence (purple), text mining evidence extracted from published literature (yellow), curated database evidence (light blue) and co-expression evidence (black). All underlying evidence can be inspected in dedicated viewers that are accessible from the network.

The novel LINE-1 insertion identified in LA PRA cases is believed to alter *IMPG2* expression in the retina and, as IMPG2 is a structural component of the IPM, probably affects the structure and/or

function of the IPM surrounding photoreceptor cells. Consequently, visual processes such as photoreceptor shedding and RPE phagocytosis may be compromised, resulting in photoreceptor degeneration over time.

BBS2

BBS2 is one of eight proteins that comprise the BBSome, a protein complex required for ciliogenesis. Using the STRING v11.0 web server database to display putative functional protein interactions suggests BBS2 interacts with other proteins within the BBSome (Figure 6.3) (Szklarczyk et al, 2019).

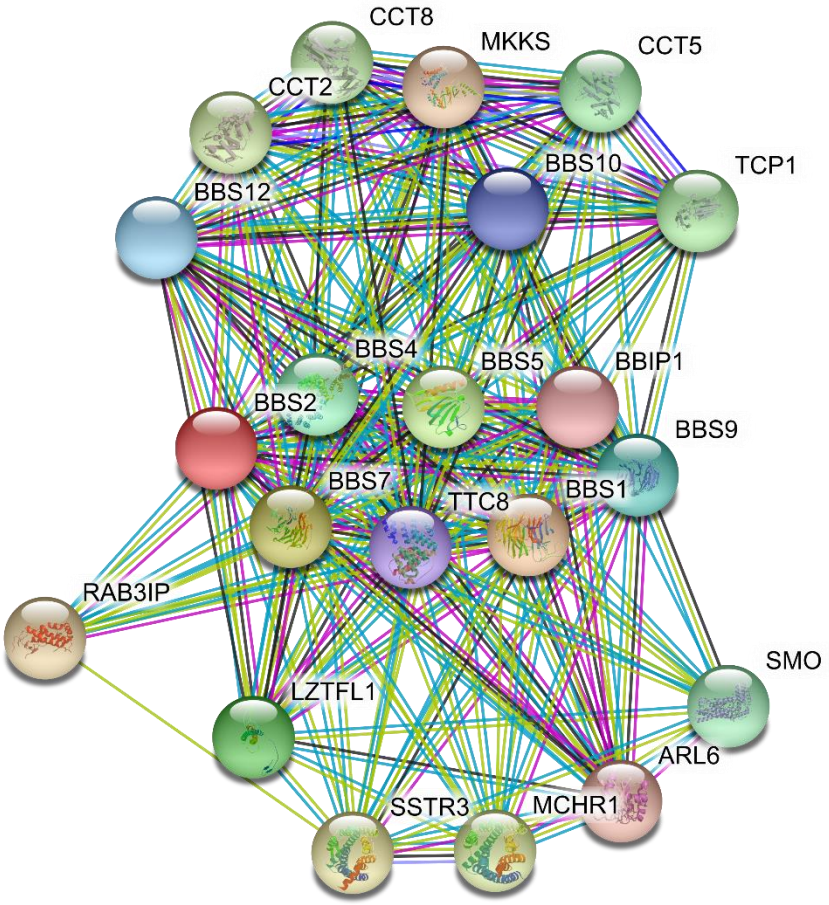


Figure 6.3: Human BBS2 protein interaction map in STRING v11.0 (Szklarczyk et al, 2019) showing the top 20 interactors with a medium confidence score (0.400). Line colour indicates the types of evidence used in predicting the associations: cooccurrence evidence (blue), experimental evidence (purple), text mining evidence extracted from published literature (yellow), curated database evidence (light blue) and co-expression evidence (black). All underlying evidence can be inspected in dedicated viewers that are accessible from the network.

Ciliogenesis occurs in photoreceptors, where connecting cilium link the outer and inner segments of photoreceptor cells. BBS proteins control formation of the primary cilia by facilitating the movement of cargo between plasma and ciliary membranes, a process that occurs across various cell types where primary cilia are present, including the retina and kidneys. Retinal photoreceptor

primary cilia and outer segments are essential for phototransduction and maturation of the RPE, suggesting the *BBS2* SNV present in SS dogs impairs retinal function through dysfunction of the phototransduction cascade. Potential renal defects exhibited in affected SS dogs may be explained by the presence of ciliogenesis in the kidney, with the *BBS2* SNV affecting transcripts encoding components of primary cilia. Another notable feature worth considering as a component of the SS *BBS2* phenotype is an increase in food motivation, suggesting that disruption of *BBS2* leads to a defect in the transport of leptin receptor to the plasma membrane, resulting in leptin resistance and obesity as demonstrated in BBS knockout mouse models (Davis et al, 2007; Rahmouni et al, 2008).

6.2 Choice of genetic applications

GWAS is a powerful tool and has, up until recently, been the method of choice in canine genetic studies of Mendelian diseases, provided DNA can be collected from enough cases and controls of the same breed and with a robust phenotype defined. In canine studies of single-gene Mendelian disorders, a GWAS approach can succeed with a modest number of samples (8 cases and 16 controls) and as few as 15,000 evenly spaced SNP markers (Lindblad-Toh et al, 2005); a fraction of the sample and marker numbers required for human GWAS. This is largely due to dog breeds existing as genetically isolated populations with relative genetic homogeneity: long stretches of LD and long haplotype blocks are present in the canine genome due to the generation of discreet dog breeds being a recent event; within the past few hundred years. Cases and controls for use in these genetic studies should ideally be determined by qualified veterinary surgeons to ensure well-defined robust sample sets. One of the key features of GWAS is that it is hypothesis free, so can uncover novel biology of disease. A benefit of GWAS mapping is that a region associated with disease can be highlighted and a candidate region can be defined, ultimately narrowing the search from genome-wide to a manageable region sometimes just Kb in size. This limits candidates to a small number of genes to manually interrogate within a region with the application of further genetic technologies such as WES or WGS for fine-mapping to pinpoint a causal variant. For the LA study, DNA from a sufficient number of cases and controls was obtained. For each individual, a clinical diagnosis was provided by a veterinary ophthalmologist to support each DNA sample, therefore providing a suitable case-control set for a GWAS to be conducted.

When the number of cases available for research is limited, such as for a newly emerging disease in a breed, a GWAS approach is not suitable. In these instances, WES or WGS can be conducted using a single case and analysing against control genomes. Using a hypothesis that the disease is private to the breed under investigation, WGS analysis can be made more powerful by using genomes of dogs affected with other diseases as control dogs to help find mutations associated with different disease in other breeds. For example, an epilepsy case can be utilised as a control in a PRA research study. The power of using multiple genomes of different breeds means that

filtering analysis extracts benign variation when looking for private high impact variants. In contrast to GWAS, knowledge of a disease-associated region is unavailable in WGS data analysis (unless GWAS SNP positions are extracted from multiple WGS data to perform GWAS analysis) as no mapping experiment has been performed. As shown with the GS and SS studies, only a small number of affected dogs were submitted to the AHT research DNA bank and this prohibited a GWAS approach. This resulted in WGS analysis with a small number of cases and related unaffected family members, using filtering steps based on the severity of the variant on the protein implicated and pedigree information to distinguish segregation, ranking variants based on their candidacy for a retinal condition.

6.3 Limitations

Phenotyping

Prior to any study of disease, individual phenotypes must be defined by a clinician. For canine ocular disease, this means a veterinary ophthalmologist must examine the individual to determine a diagnosis. However, qualifications and experience of ophthalmologists can vary (certificates and diplomas). In addition, as with many research studies, many individuals within a study cohort are typically examined by multiple ophthalmologists as opposed to a single clinician performing all ophthalmoscopic examinations. The clinical information collected for each individual is also varied, where some cases submitted to research include supporting reports with ophthalmic observations, fundus information or photographs, and others simply mention a PRA-diagnosis. For this reason, in many investigations of hereditary ocular disease an individual is considered a robust case when examined by a certified veterinary ophthalmologist, a European diploma holder or a BVA/KC/ISDS eye panellist, where evidence is provided to support the sample with detailed information and diagrams of fundus abnormalities, where possible. Researchers rely on an accurate diagnosis of PRA, especially where a single case is being analysed, for example using a WGS approach, in order to minimise costs and time in identifying a causal PRA-variant. The presence of PRA phenocopies that display similar clinical signs to the disease can sometimes result in a misdiagnosis. This includes SARD, as exemplified in the LA cohort, which presents ophthalmoscopic signs of retinal degeneration that may be mistaken for PRA. Therefore, it is important to consider PRA-phenocopies during any molecular or clinical studies. Where multiple cases are collected for a larger study, such as a GWAS, including individuals with ambiguous phenotypes may reduce the power of a sample set but in turn defining an unnecessarily strict phenotype can impede sample collection.

Age of onset

For PRA, the age of onset is generally not defined by the age of diagnosis. Determining the exact age of onset can be difficult to assess when the disease is progressive, and owners may remain unaware their dog is affected until visual impairment becomes severe. Many dogs can adapt to

vision loss and compensate with their highly sensitive sense of smell and behavioural changes, especially if they share their environment with other dogs or rarely go out of familiar surroundings. The lack of detailed ERG evaluation in dogs, which would detect early changes in rod and cone function before the dog started showing any behavioural changes, often leaves ophthalmoscopic examination as the extent of the diagnosis.

Confirmation of causality

The absence of tissue, specifically retina, from affected dogs is a common issue in canine PRA investigations, where the majority of affected dogs rarely undergo enucleation on welfare grounds and are typically lost to research follow-up. This limits histology, gene expression or protein analysis to explore the consequences of the mutations identified. As a result, bioinformatics tools are used to predict variant consequence. *In silico* tools such as SIFT, PolyPhen and MutationTaster classify SNVs according to negative, neutral or positive effects on protein structure or function. Although these prediction algorithms can be beneficial in shortlisting variants, these tools do not confirm the nature of the variant. Comparison with human biology and model organisms can also provide supporting evidence of confirmation of causality.

Structural variation

WGS analysis used here implements filtration steps to search for SNVs or small INDELs. More complex mutations such as inversions, translocations or transposable elements are difficult to identify (as discussed further below) and therefore are not included in the current AHT WGS analysis pipeline. For this reason, large structural variation cannot be fully excluded in the WGS studies undertaken in this thesis. Detection of structural variants from the use of short-read sequencing technologies in WGS is onerous. Poor breakpoint resolution is achieved in repetitive regions of the genome and the identification of structural variation relies on a combinatorial approach using paired-end mapping and read depth analysis on a genome-wide scale. As structural variants are extremely diverse in type and size, data analysis does not always produce a manageable list to follow up. To overcome these challenges, long-read sequencing technologies, such as PacBio or Oxford Nanopore sequencing, could be used alone or in conjunction with short-read Illumina sequencing to mediate both SNV/INDEL detection and complex structural variant identification. Such methods have been utilised in human studies into Mendelian disease, including a homozygous deletion-inversion-deletion complex structural variant in the *CEP78* gene in a patient with cone-rod dystrophy identified using short-read WGS (Sanchis-Juan et al, 2018). Due to technical challenges in identification, complex structural variants are often not considered during genetic analyses, although they should be considered in future studies, especially for unresolved PRA cases. Although the majority of PRA-associated variants identified to date are within coding regions, the LINE-1 insertion within the *IMPG2* promoter identified in this study demonstrates that non-coding variants cannot be excluded as causal in PRA studies. More complex

mutations may be causing PRA in unresolved cases, where when a WGS approach is used without GWAS, such variants have not been possible to detect with current filtering steps. More efforts must be made to validate and implement structural variant tools in analysis pipelines for canine genetic studies in order to facilitate structural variant detection in rare diseases.

Non-coding variants

Until recently, all reported PRA-associated variants lie within protein-coding regions of genes (or promoter regions). The only report-to-date of a regulatory non-coding variant implicated in a canine PRA is an intronic SNV in the transcription factor encoding the ortholog of human immunodeficiency virus type I enhancer binding protein 3 gene (*HIVEP3*) in Miniature Schnauzer dogs (Kaukonen et al, 2020). Defining the functional consequences of non-coding variants remains a challenge, however such variants have been associated with retinal diseases in humans; including deep intronic variants that may cause aberrant splicing events, create cryptic splice sites, or disrupt regulatory elements (Bax et al, 2015; Fadaie et al, 2019; Webb et al, 2012). In the studies described in this thesis candidate variants were identified, which segregated with the disease, and evidence of their causality is described. Thus, further analysis of non-coding variants was not performed. However, in future PRA studies where initial analysis of WGS data do not highlight candidate variants, non-coding variants should also be considered.

6.4 Future directions

In each investigation, the PRA-associated variant identified was genotyped across dogs of other breeds. Both the LA and SS variants were absent from the individuals tested from other breeds, however more extensive genotyping from larger breed cohorts would be required to fully exclude the presence of these variants in other breeds, in particular those with shared ancestry. The GS *NECAP1* SNV was found in Miniature Long-Haired Dachshunds and Spitz varieties, as well as mixed-breed dogs sharing a proportion of their ancestry to these breeds. Studying the haplotype around the SNV in the GS, Miniature Long-Haired Dachshund and a Giant Spitz showed a shared haplotype, with a shorter haplotype in the Miniature Long-Haired Dachshund breed, suggesting the variant is inherited identically by descent. Breed histories suggest these breeds are probably of German ancestry, proposing that the SNV arose from a common ancestral chromosome containing the variant in a single ancestral founder dog predating the formation of each breed. There are examples of pathogenic mutations in dogs that are inherited by descent across multiple related breeds, including PRCD-PRA (Goldstein et al, 2006; Zangerl et al, 2006) and CEA (Parker et al, 2007). Further study of the breeds found to share the *NECAP1* SNV is required to determine genotype-phenotype correlations and establish whether the mutation is causal across these additional breeds.

The absence of retinal tissue from any individuals in these studies prevented functional follow-up work to be performed. For this thesis, RNA-seq analysis would have been favourable for all breeds

investigated. RNA-seq data from PRA cases and controls of each breed investigated may have facilitated further characterisation of each candidate gene and define whether retinal-specific transcripts were present and thus disrupted in PRA-affected individuals. This would be particularly advantageous in the GS study where *NECAP1*, a novel retinal candidate gene, is implicated. Alongside previously published transcriptome studies in the dog, analysis of RNA-seq data in our laboratory from retinal tissue from a PRA non-affected dog demonstrated that *NECAP1* is expressed in the normal canine retina. Comparison of this with retinal tissue from a PRA-affected GS would establish if *NECAP1* expression is altered in affected dogs. Retinal transcripts could also be investigated to determine if there are *NECAP1* retinal-specific transcripts in the dog compared to humans. As *NECAP1* mutations in humans are associated with a neurological disease, comparison of RNA-seq data generated from brain tissue and retinal tissue in affected and non-affected dogs, and comparison to the human transcripts, could provide further insight into why a retinal phenotype is associated with *NECAP1* in the dog, but not directly in humans. In addition to this, retinal cell lines from a LA PRA case and a LA control dog would enable a luciferase assay to be optimised and performed, determining the impact the LINE-1 insertion on *IMPG2* expression.

Conducting a clinical study into the *BBS2* phenotype may improve our understanding of the clinical manifestations observed, providing further disease characterisation and to determine the presence of phenotypic variability in the disease onset, progression or severity of secondary features across individuals. Ethical concerns must be considered prior to recruiting *BBS2* homozygotes for a clinical study, for example establishing a donor programme to harvest multiple tissue types following euthanasia, or establishing a research colony to control environmental factors such as age and diet. This would enable histological experiments and RNA analyses to be performed, for example to investigate retinal and renal defects, to investigate molecular pathways involved in the disease. Another application to consider with the availability of tissue samples are immunohistochemistry experiments, ideally comparing tissue sections from a normal canine retina and a PRA-affected retina. Immunohistochemistry utilises monoclonal and polyclonal antibodies to detect specific antigens in sections of a tissue of interest and requires isolated, sectioned tissue and an appropriate antibody. Establishing where in the retina a gene is expressed, for example if primarily expressed in photoreceptor cells, would provide additional evidence for disease causality. In addition to further exploration of *BBS2* phenotypes in SS, this application could be considered for future work investigating the *NECAP1* variant in GS dogs, subject to tissue availability.

Although beyond the scope of this thesis, hypothetically animal models could be used to provide further understanding of the pathogenesis of *NECAP1* in ocular disease. Utilisation of mouse or zebrafish models as a means of *in vivo* study of gene function have been used to further elucidate the functional effects in genes of interest (Stiebel-Kalish et al, 2012). For *NECAP1*, no knockout

animal models have been reported so far, however implementing this approach by *NECAP1* inactivation may provide insight into mouse or zebrafish phenotypes. Furthermore, studies using zebrafish models, specifically in ocular disease studies, have proven successful using the CRISPR/Cas9 system to target and repair specific mutations in zebrafish (Irion et al, 2014). Showing similarities in eye development, retinal structure and gene orthology to humans, the zebrafish is a prominent model organism in the study of human ocular disease (Richardson et al, 2017). Such an application could be implemented to assess the *NECAP1* SNV in a zebrafish model, using CRISPR/Cas9 to introduce the mutation and conduct *in vivo* experiments of the impact of the *NECAP1* c.544G>A SNV on zebrafish retinal morphology. Finally, these studies can provide a framework for further exploration of each variant and its involvement in disease pathogenesis. Harvesting canine induced pluripotent skin cells from affected dogs, differentiating into RPE and photoreceptor cells and performing *in situ* gene correction using CRISPR/Cas9 of each mutation would provide further understanding of whether these molecular and cellular phenotypes can be rescued.

Finally, there are still dog breeds suffering from PRA for which a molecular diagnosis has not yet been found. DNA from PRA cases of other breeds has been collected and stored at the AHT, and known retinal mutations have been excluded as causal. Some of these breeds include in the Irish Red and White Setter, English Shepherd, Miniature Long-Haired Dachshund, Tibetan Spaniel and Tibetan Terrier, and sample numbers range from single to multiple cases per breed. This emphasises the genetic heterogeneity of PRA across and within breeds. Discussions with individual Breed Clubs, their Breed Health Coordinators and veterinary ophthalmologists to define clear phenotypes and collect additional robust cases and controls, along with pedigree information, will help decide which genetic techniques will be most appropriate to pursue genetic investigations for breeds with unknown molecular diagnoses.

6.5 Conclusions

This thesis describes three novel variants in three dog breeds and provides supportive evidence for their causality in canine PRA. Within each breed, ophthalmoscopic evaluation by veterinary ophthalmologists provided robust cases and controls for each genetic study, highlighting the importance of obtaining clear clinical information for canine genetic studies. As demonstrated, this is particularly significant in the absence of appropriate tissue to conduct further functional analysis for a variant of interest, relying on genotyping large breed cohorts and determining allele frequencies to provide additional evidence that a variant is segregating with a disease. This, along with the use of bioinformatics *in silico* prediction tools, provided evidence that each variant identified is associated with disease, in the breeds implicated. As a result, diagnostic DNA tests have been developed to use as a breeding tool aiming to reduce the frequency of, and eventually eradicate, the mutation from each breeding population within a reasonable timeframe. With many

unresolved cases for canine PRA, human IRD studies have proven significant in facilitating the identification of canine PRA-associated genes. As illustrated, novel gene discovery in canine PRA studies can also present genes for consideration in retinal degeneration in humans and other species to identify disease-causing variants. Finally, this thesis contributes three novel canine PRA-associated mutations to the community, highlighting the potential applications and challenges in elucidating the genetic basis of canine PRAs.

Appendices

Appendix 1. Genes from the RetNet database associated with human inherited retinal diseases.

Retinitis pigmentosa, autosomal dominant	Retinitis pigmentosa, autosomal recessive		Retinitis pigmentosa, X-linked	Leber congenital amaurosis, autosomal dominant	Leber congenital amaurosis, autosomal recessive	Bardet-Biedl syndrome, autosomal recessive
ADIPOR1	*ABCA4*	*MAK*	*OFD1*	*CRX*	*AIPL1*	*ADIPOR1*
ARL3	*AGBL5*	*MERTK*	*RP2*	*IMPDH1*	*CABP4*	*ARL6*
BEST1	*AHR*	*MVK*	*RPGR*	*OTX2*	*CCT2*	*BBIP1*
CA4	*ARHGEF18*	*NEK2*			*CEP290*	*BBS1*
CRX	*ARL6*	*NEUROD1*			*CLUAP1*	*BBS2*
FSCN2	*ARL2BP*	*NR2E3*			*CRB1*	*BBS4*
GUCA1B	*BBS1*	*NRL*			*CRX*	*BBS5*
HK1	*BBS2*	*PDE6A*			*DTHD1*	*BBS7*
IMPDH1	*BEST1*	*PDE6B*			*GDF6*	*BBS9*
IMPG1	*C2orf71*	*PDE6G*			*GUCY2D*	*BBS10*
KLHL7	*C8orf37*	*POMGNT1*			*IFT140*	*BBS12*
NR2E3	*CERKL*	*PRCD*			*IQCB1*	*C8orf37*
NRL	*CLCC1*	*PROM1*			*KCNJ13*	*CEP19*
PRPF3	*CLRN1*	*RBP3*			*LCA5*	*CEP290*
PRPF4	*CNGA1*	*REEP6*			*LRAT*	*IFT172*
PRPF6	*CNGB1*	*RGR*			*NMNAT1*	*IFT27*
PRPF8	*CRB1*	*RHO*			*PRPH2*	*INPP5E*
PRPF31	*CYP4V2*	*RLBP1*			*RD3*	*KCNJ13*

Appendix 1 (continued).

Retinitis pigmentosa, autosomal dominant	**Retinitis pigmentosa, autosomal recessive**		**Retinitis pigmentosa, X-linked**	**Leber congenital amaurosis, autosomal dominant**	**Leber congenital amaurosis, autosomal recessive**	**Bardet-Biedl syndrome, autosomal recessive**
PRPH2	*DHDDS*	*RP1*			*RDH12*	*LZTFL1*
RDH12	*DHX38*	*RP1L1*			*RPE65*	*MKKS*
RHO	*EMC1*	*RPE65*			*RPGRIP1*	*MKS1*
ROM1	*EYS*	*SAG*			*SPATA7*	*NPHP1*
RP1	*FAM161A*	*SAMD11*			*TULP1*	*SDCCAG8*
RP9	*GPR125*	*SLC7A14*				*TRIM32*
RPE65	*HGSNAT*	*SPATA7*				*TTC8*
SAG	*IDH3B*	*TRNT1*				
SEMA4A	*IFT140*	*TTC8*				
SNRNP200	*IFT172*	*TULP1*				
SPP2	*IMPG2*	*USH2A*				
TOPORS	*KIAA1549*	*ZNF408*				
	KIZ	*ZNF513*				

Appendix 2. Mutations associated with canine inherited retinal disease (IRD) and those excluded in the laboratory.

	IRD/ Abbreviation	Inheritance	Gene/ Locus	References	CanFam3 Position	Mutation	Breeds identified	Excluded by laboratory screening
				Cone-rod degenerations (CRDs)				
1	Cone-rod dystrophy 1 (CORD1)	Autosomal recessive	*RPGRIP1*	(Mellersh et al, 2006)	chr15:18,331,995	44-bp insertion	Miniature Long-Haired Dachshund	LA, GS, SS
2	EORD modifier	Autosomal recessive	*MAP9*	(Forman et al, 2016)	chr15:52,907,604-52,908,043	Deletion	Miniature Long-Haired Dachshund	LA, GS, SS
3	Cone-rod dystrophy 3 (CRD3)	Autosomal recessive	*ADAM9*	(Kropatsch et al, 2010)	Approx. chr16:26446610-26469830	Deletion of exons 15 and 16	Glen of Imaal Terrier	LA, GS, SS
4	Cone-rod dystrophy (CRD)	Autosomal recessive	*NPHP4*	(Wiik et al, 2008)	chr5:59,913,209	180-bp deletion	Standard Wire-haired Dachshund	LA, GS, SS
5	Cone-rod dystrophy 1 (crd1)	Autosomal recessive	*PDE6B*	(Goldstein et al, 2013b)	chr3:91747728-91747730	3-bp del (AAC)	American Staffordshire Terrier	N/A

Appendix 2 (continued).

	IRD/ Abbreviation	Inheritance	Gene/ Locus	References	CanFam3 Position	Mutation	Breeds identified	Excluded by laboratory screening
				Progressive retinal atrophies (PRAs)				
6	Cone-rod dystrophy 2 (crd2)	Autosomal recessive	*IQCB1(NPHP5)*	(Goldstein et al, 2013b)	chr33:25078909 -25078910	1-bp insertion (C)	American Pitbull Terrier	N/A
7	Rod-cone dystrophy 1 (RCD1a)	Autosomal recessive	*PDE6B*	(Dekomien et al, 2000)	chr3:91747685-91747686	8bp insertion (TGAAGTCC)	Sloughi	LA, GS, SS
8	Rod-cone dystrophy 1 (RCD1)	Autosomal recessive	*PDE6B*	(Clements et al, 1993; Suber et al, 1993)	chr3:91747714	Missense SNV (G/A)	Irish Setter	LA, GS, SS
9	Rod cone dysplasia type 2 (rcd2)	Autosomal recessive	*C1orf36 (RD3)*	(Kukekova et al, 2009)	Approx. chr7:9874987:9876615	22-bp insertion	Collie	N/A
10	Rod-cone dystrophy 3 (RCD3)	Autosomal recessive	*PDE6A*	(Petersen-Jones et al, 1999)	chr4:59145361	1-bp deletion (A)	Cardigan Welsh Corgi	LA, GS, SS

Appendix 2 (continued).

	IRD/ Abbreviation	Inheritance	Gene/ Locus	References	CanFam3 Position	Mutation	Breeds identified	Excluded by laboratory screening
11	Rod-cone degeneration 4 (RCD4)	Autosomal recessive	*C2ORF71*	(Downs et al, 2013)	chr17:22907389-22907390	1-bp insertion (C)	Gordon Setter, Irish Setter, English Setter & Tibetan Terrier	LA, GS, SS
12	Progressive retinal atrophy 3 (PRA3)	Autosomal recessive	*FAM161A*	(Downs & Mellersh, 2014)	chr10:61,822,335-61,822,372	SINE insertion	Tibetan Spaniel & Tibetan Terrier	LA, GS, SS
13	Early retinal degeneration (ERD)	Autosomal recessive	*STK38L*	(Goldstein et al, 2010)	chr27:20,447,816	4-bp deletion in exon 3, SINE insertion in exon 4	Norwegian Elkhound	LA, GS, SS
14	Generalized PRA (GPRA)	Autosomal recessive	*CCDC66*	(Dekomien et al, 2010)	chr20:33,745,446-33,745,447	1-bp insertion (A)	Schapendoes	LA, GS, SS
15	Progressive retinal atrophy (PRA)	Autosomal recessive	*CNGB1*	(Ahonen et al, 2013)	chr2:58622673; chr2:58622676-58622677	1-bp deletion (A); 6-bp insertion (AGCTAC)	Papillion & Phalènes	LA, GS, SS

Appendix 2 (continued).

	IRD/ Abbreviation	Inheritance	Gene/ Locus	References	CanFam3 Position	Mutation	Breeds identified	Excluded by laboratory screening
16	Progressive rod-cone degeneration (PRCD)	Autosomal recessive	*PRCD*	(Goldstein et al, 2006; Zangerl et al, 2006)	chr9:4188663	Missense SNV (G/A)	72 breeds (i)	LA, GS, SS
17	Autosomal dominant PRA (ADPRA)	Autosomal dominant	*RHO*	(Kijas et al, 2002; Kijas et al, 2003)	chr20:5,637,394	Missense SNV (C/G)	English Mastiff & Bull Mastiff	LA, GS, SS
18	Golden Retriever PRA 1 (GR_PRA_1)	Autosomal recessive	*SLC4A3*	(Downs et al, 2011)	chr37:26145747 -26145748	1-bp insertion (C ins)	Golden Retriever	LA, GS, SS
19	Golden Retriever PRA 2 (GR_PRA_2)	Autosomal recessive	*TTC8*	(Downs et al, 2014)	chr8:60090186	1-bp deletion (A)	Golden Retriever and Labrador Retriever	LA, GS, SS
20	Progressive retinal atrophy (PRA)	Autosomal recessive	*CNGA1*	(Wiik et al, 2015)	chr13:43,831,896-43,831,900	4-bp deletion	Shetland Sheepdogs	LA, GS, SS
21	X-linked PRA (XLPRA1)	X-linked recessive	*RPGR*	(Zhang et al, 2002)	Approx. chrX:33,077,524-33077426	5-bp deletion (GAGAA)	Siberian Husky & Samoyed	N/A

Appendix 2 (continued).

	IRD/ Abbreviation	Inheritance	Gene/ Locus	References	CanFam3 Position	Mutation	Breeds identified	Excluded by laboratory screening
22	X-linked PRA (XLPRA2)	X-linked recessive	*RPGR*	(Zhang et al, 2002)	Approx. chrX:33,077,524-33077426	2-bp del (GA)	Mongrel dogs (mixed breeds); Miniature Schnauzer	N/A
23	X-linked PRA	X-linked recessive	*RPGR*	(Kropatsch et al, 2016)		Large deletion	Weimaraner	N/A
24	Canine Stargardt Disease	Autosomal recessive	*ABCA4*	(Makelainen et al, 2019)	chr6:55146549	1-bp insertion (C)	Labrador Retriever	SS
25	PRA	Autosomal recessive	*SAG*	(Goldstein et al, 2013a)	chr25:44843440	Missense SNV (T/C)	Basenji	N/A
26	PRA	Autosomal recessive	*PPT1*	(Murgiano et al, 2018)		Structural variant	Miniature Schnauzer	N/A
27	PRA	Autosomal recessive	*HIVEP3*	(Kaukonen et al, 2020)	chr15: 1432293	Intronic SNV (G/A)	Miniature Schnauzer	N/A

Appendix 2 (continued).

	IRD/ Abbreviation	Inheritance	Gene/ Locus	References	CanFam3 Position	Mutation	Breeds identified	Excluded by laboratory screening
Other								
N/A	Photoreceptor Dysplasia	Autosomal recessive	PDC	(Zhang et al, 1998)	chr7:19,499,691	Missense SNV (C/G)	Miniature Schnauzer however now not believed to be a causal mutation	LA, GS, SS
26	Retinopathy	Autosomal recessive	MERTK	(Everson et al, 2017)	chr17:36338043 -36338057	LINE-1 insertion	Swedish Vallhund	SS
27	Achromatopsia (ACHM)	Autosomal recessive	CNGB3	(Yeh et al, 2013)	chr29:32837065	Missense SNV (G/A)	Alaskan Malamute, German Short-Haired Pointer	LA, GS, SS
28	Achromatopsia (ACHM)	Autosomal recessive	CNGA3	(Tanaka et al, 2015)	chr10:44234861	Missense SNV (C/T)	German Shepherd	LA, GS, SS
29	Achromatopsia (ACHM)	Autosomal recessive	CNGA3	(Tanaka et al, 2015)	chr10:44234200	3-bp del (TGG)	Labrador Retriever	LA, GS, SS
30	Canine LCA (cLCA)	Autosomal recessive	RPE65	(Veske et al, 1999)	chr6:76893207-76893210	4bp deletion (AAGA)	Briard	LA, GS, SS

Appendix 2 (continued).

	IRD/ Abbreviation	Inheritance	Gene/ Locus	References	CanFam3 Position	Mutation	Breeds identified	Excluded by laboratory screening
31	Canine multifocal retinopathy 1 (cmr1)	Autosomal recessive	*BEST1*	(Guziewicz et al, 2007)	chr18:54478586	Missense SNV (C/G)	Great Pyrenees, English Mastiff & Bullmastiff	LA, GS, SS
32	Canine multifocal retinopathy 2 (cmr2)	Autosomal recessive	*BEST1*	(Guziewicz et al, 2007)	chr18:54,476,143	Missense SNV (G/A)	Coton de Tulear	LA, GS, SS
33	Canine multifocal retinopathy 3 (cmr3)	Autosomal recessive	*BEST1*	(Zangerl et al, 2010)	chr18:54,470,587	Deletion (C)	Lapponian Herder	LA, GS, SS

(i) American Cocker Spaniel, American Eskimo Dog, American Hairless Terrier, Australian Cattle Dog, Australian Shepherd, Australian Silky Terrier, Australian Stumpy Tail Cattle Dog, Barbet, Black Russian Terrier, Bolognese, Bolonka, Chesapeake Bay Retriever, Chihuahua, Chihuahua – Longhaired, Chihuahua - Smooth-haired, Chinese Crested Dog, Coton de Tulear, English Cocker Spaniel, Entlebucher Mountain Dog, Field Spaniel, Finnish Lapphund, Giant Schnauzer, Giant Schnauzer – Black, Giant Schnauzer - Pepper and salt, Golden Labrador Retriever, Golden Retriever, Goldendoodle, Karelian Bear Dog, Kuvasz, Labrador Retriever, Lagotto Romagnolo - Romagna Water Dog, Lancashire Heeler, Lapponian Herder, Manchester Terrier – Toy, Miniature American Shepherd, Mixed breed, Norrbottenspitz, Norwegian Elkhound, Black, Norwegian Elkhound, Grey, Nova Scotia Duck Tolling Retriever, Plott, Poodle (AKC), Poodle - Medium size (FCI size standard) - Black, brown and white, Poodle - Medium size (FCI size standard) - Grey, apricot and red, Poodle - Miniature (AKC size standard), Poodle - Miniature (FCI size standard) - Black, brown and white, Poodle - Miniature (FCI size standard) - Grey, apricot and red, Poodle - Standard (AKC size standard), Poodle - Standard (FCI size standard} - Black, brown and white, Poodle - Standard (FCI size standard) - Grey, apricot and red, Poodle - Toy (AKC size standard), Poodle - Toy (FCI size standard), Portuguese Podengo, Portuguese Podengo – Miniature, Portuguese Podengo - Smooth-haired Miniature, Portuguese Podengo - Wirehaired Miniature, Portuguese Water Dog, Rat Terrier, Schipperke, Spanish Water Dog, Swedish Elkhound, Swedish Lapphund, Xoloitzcuintle, Yorkshire Terrier

Appendix 3. The BVA/KC/ISDS eye scheme listing canine inherited eye diseases and the breeds advised to have annual eye examinations (January 2020).

canine health schemes

BVA/KC/ISDS Eye Scheme – Inherited Eye Disease: Breeds affected and type of disease (formally known as Schedule A)

Alphabetical list of breeds and their eye conditions for certification under the Inherited Eye Diseases Status section of the Certificate of Examination

1. Alaskan Malamute – **HC**
2. Australian Cattle Dog – **PRA**
3. Australian Shepherd – **HC**
4. Basset Hound - **G, POAG**
5. Bedlington Terrier – **TRD**
6. Belgian Shepherd Dog (all varieties) – **HC**
7. Bichon Frise – **HC**
8. Border Collie - **CEA, RPED, PLL, G**
9. Boston Terrier - **HC (two forms)**
10. Briard – **RPED**
11. Bull Terrier (Miniature) – **PLL**
12. Cavalier King Charles Spaniel - **MRD, HC**
13. Collie (Rough) - **CEA, PRA, RPED**
14. Collie (Smooth) - **CEA, RPED**
15. Dachshund (Miniature Long-Haired) – **PRA**
16. Dandie Dinmont – **G**
17. Dobermann – **PHPV**
18. Finnish Lapphund – **PRA**
19. Fox Terrier (Smooth) – **PLL**
20. Fox Terrier (Wire) – **PLL**
21. French Bulldog – **HC**
22. German Shepherd Dog – **HC**
23. Giant Schnauzer – **HC**
24. Glen of Imaal Terrier – **PRA**
25. Gordon Setter – **PRA**
26. Hungarian Puli – **MRD**
27. Hungarian Vizsla - **G**
28. Irish Red and White Setter – **HC**
29. Irish Setter – **PRA**
30. Irish Wolfhound – **PRA**
31. Japanese Shiba Inu – **G**
32. Lancashire Heeler - **CEA, PLL**
33. Large Munsterlander – **HC**
34. Leonberger – **G, HC**
35. Lhasa Apso – **PRA**
36. Miniature Schnauzer - **CHC, PRA, HC**
37. Norwegian Buhund – **HC**
38. Norwegian Elkhound – **PRA**
39. Old English Sheepdog – **HC**
40. Papillon - **PRA**
41. Parson Russell Terrier – **PLL**
42. Petit Basset Griffon Vendeen – **POAG**
43. Poodle (Miniature) – **PRA**
44. Poodle (Standard) – **HC**
45. Poodle (Toy) – **PRA**
46. Retriever (Chesapeake Bay) - **PRA, HC**
47. Retriever (Flat Coated) – **G**
48. Retriever (Golden) - **MRD, PRA, RPED, HC, G**
49. Retriever (Labrador) - **MRD, TRD, PRA, RPED, HC**
50. Retriever (Nova Scotia Duck Tolling) – **PRA**
51. Rottweiler – **MRD**
52. Sealyham Terrier - **TRD, PLL**
53. Shar Pei – **POAG**
54. Shetland Sheepdog - **CEA, RPED**
55. Siberian Husky - **G, HC**
56. Spaniel (American Cocker) - **MRD, G, PRA, HC**
57. Spaniel (Cocker) - **G, PRA, RPED**
58. Spaniel (English Springer) - **MRD, G, PRA, RPED**
59. Spaniel (Welsh Springer) - **G, HC**
60. Spanish Water Dog – **G**
61. Staffordshire Bull Terrier - **PHPV, HC**
62. Swedish Vallhund **BR, PRA**
63. Tibetan Spaniel – **PRA**
64. Tibetan Terrier - **PRA, PLL**
65. Welsh Corgi (Cardigan) - **PRA, RPED**

Inherited Eye Disease Status - Key to abbreviations:

G = Goniodysgenesis/Primary Glaucoma	**POAG** = Primary Open Angle Glaucoma	**PLL** = Primary Lens Luxation	**CHC** = Congenital Hereditary Cataract	**OM** = Ocular Melanosis
HC = Hereditary Cataract	**PHPV** = Persistent Hyperplastic Primary Vitreous	**CEA** = Collie Eye Anomaly	**MRD** = Multifocal Retinal Dysplasia	**BR** = Breed Specific Retinopathy
TRD = Total Retinal Dysplasia	**RPED** = Retinal Pigment Epithelial Dystrophy (formerly Central Progressive Retinal Atrophy = CPRA)	**PRA** = Progressive Retinal Atrophy		

The **British Veterinary Association** and **the Kennel Club** — working together for excellence in canine health

BVA British Veterinary Association

THE KENNEL CLUB *Making a difference for dogs*

January 2020 | 1

THE INTERNATIONAL SHEEP DOG SOCIETY

Appendix 4. All primers and assay details.

Table A1: PCR primers used for sequencing amplicons to exclude known canine retinal mutations as causal in the PRA- affected Giant Schnauzer (GS), Lhasa Apso (LA) and Shetland Sheepdog (SS) dogs under investigation. (B, C) PCR primers used for genotyping seven known canine retinal mutations to exclude in the breeds under investigation (by PCR followed by AFLP analysis or by visualisation of PCR product on an agarose gel).

Form of retinal degeneration	Locus or Abbreviation	Gene involved	Primer sequences	References
(A) Pooled for multiplex PCR for amplicon sequencing				
Cone degeneration	CD	*CNGB3*	F-CCCACTTACCATTATGTCTCCTC; R-TCAAACACCAGACAACACACA	(Yeh et al, 2013)
Canine achromatopsia	ACHM	*CNGA3*	SNP_F-AGTGGGCGTCCTAATCTTTG; SNP_R-CGGAATTGCATGTACTGCTT	(Tanaka et al, 2015)
Canine achromatopsia	ACHM	*CNGA3*	del_F-CCTGATCGACGAGGACGTG; del_R-CGTTGTACTCGGCCAGGAG	(Tanaka et al, 2015)
Congenital stationary night blindness/retinal dystrophy	CSNB	*RPE65*	F-GACAATGCCCTTGTTAACGTC; R-AGCATTTCGTCCTACCTGCTT	(Veske et al, 1999)
Canine multifocal retinopathy	CMR1	*BEST1/ VMD2*	F-CGTCACCTACTCAAGCCAAGT; R-TAGCTGAGCAGGAAGATGAGG	(Guziewicz et al, 2007)
Canine multifocal retinopathy	CMR2	*BEST1/ VMD2*	F-GGCCCTCCCCATCCACTG; R-CAGGGCACCCAGAACGTGT	(Guziewicz et al, 2007)
Canine multifocal retinopathy	CMR3	*BEST1/ VMD2*	F-AAGGCCTGGGAGATTAAGGAG; R-CTGACTGTCCAGATGGGAAGA	(Zangerl et al, 2010)
Generalised PRA	GPRA	*CCDC66*	F-GCTCATACTGCTTAGATGTCTCTTG; R-CAATACTTCCAATCAGCCAAA	(Dekomien et al, 2010)
Progressive retinal atrophy	PRA	*CNGB1*	F-CCACCCAGGCTCAGCAG; R-TCAGGCAGCCCACCAAT	(Ahonen et al, 2013)

Appendix 4 (continued).

Form of retinal degeneration	Locus or Abbreviation	Gene involved	Primer sequences	References
Photoreceptor dysplasia	PD	*PDC*	F-TGTCTATTCCCAGATGAGCATT; R-AAACTCAGCTTCTGGTGCATATC	(Zhang et al, 1998)
Progressive rod-cone degeneration	PRCD	*PRCD*	F-TTTCTCCTGCAGACTCTGTCC; R-CAGCTTCTCACGGTTGGA	(Goldstein et al, 2006; Zangerl et al, 2006)
Autosomal dominant PRA	ADPRA	*RHO*	F-ACACCCGTCTTGTTGGAGAA; R-CCTCAGCAGCACTCTTAGGAC	(Kijas et al, 2002; Kijas et al, 2003)
Progressive retinal atrophy	GR_PRA1	*SLC4A3*	F-CTGCCCAGGTGAGTGCTAGA; R-CACTCCCGTTCAGCTCCAG	(Downs et al, 2011)
Progressive retinal atrophy	GR_PRA2	*TTC8*	F-TGGACTTGGCTGCCCTTT; R-CCATGTCTAAGCCCTTCACAA	(Downs et al, 2014)
Progressive retinal atrophy	PRA	*CNGA1*	F-GCAATCGAAGAACAGCCAAT; R-GCTTTCCCTTCTCTTCTAGCAT	(Wiik et al, 2015)
Rod-cone dystrophy	RCD1	*PDE6B*	F-GAAGAGATCCTGCCCATGTT; R-TGTCCTCTTGTTGCTGCTTCT	(Clements et al, 1993; Suber et al, 1993)
Rod-cone dystrophy	RCD1a	*PDE6B*	F-GAAGAGATCCTGCCCATGTT; R-TGTCCTCTTGTTGCTGCTTCT	(Dekomien et al, 2000)
Rod-cone dystrophy	RCD3	*PDE6A*	F-CAGGACTGGGTGAGGATGATA; R-ACTTGAAATACAGGGCGAGGT	(Petersen-Jones et al, 1999)
(B) Amplified fragment length polymorphism (AFLP) analysis				
Cone-rod dystrophy	CORD1	*RPGRIP1*	6FAM-GAAGAGCACATGTTGGTGAAGG; R- TGAGCTTTGTTTGCCTTTGG	(Mellersh et al, 2006)
Cone-rod dystrophy (early onset)	EORD	*MAP9*	6FAM-GTCGATGGAGGTCTCCGTAT; R- AACCTGGACCATGGGCAATA	(Forman et al, 2016)

Appendix 4 (continued).

Form of retinal degeneration	Locus or Abbreviation	Gene involved	Primer sequences	References
Rod-cone degeneration	RCD4	*C2ORF71*	6FAM-CCGAGTGCTCCCTCTGTG ; R- GGCTGCAGGCCTCGTC	(Downs et al, 2013)
Progressive retinal atrophy	PRA3	*FAM161A*	Aff_F- GGATCCCTTTATTTGATTTTAGAAAG; Nor_F- TCCCTTCCTTTTATTTGATTTTAGAAAG; R- 6FAM- CAACAAACACAACCTGAGCAA	(Downs & Mellersh, 2014)
(C) Agarose gel electrophoresis				
Cone-rod dystrophy	CRD3	*ADAM9*	Nor_F- AAGTACCTACCCTCCCGTTCA; Nor_R- GTAGGATGGAGGTCGAAGAGG; Aff_F- TTGGAACACCTAAATGCCTTG; Aff_R- GTAACCCCAGCCAACACAGTA	(Kropatsch et al, 2010)
Cone-rod dystrophy	CRD	*NPHP4*	F- AGTTCCTTCAGTTACGTCAACCAT; R- CTTCCAACACACAGACAGTACTCC	(Wiik et al, 2008)
Early retinal degeneration	ERD	*STK38L*	F- CAGCCGCACCTAAAATGTAAA; R- CTTTATTCTTGGGTGAACTCCA	(Goldstein et al, 2010)

Table A2: Primers used in Chapter 3.

Primer Name	Assay Method Used	Primer Sequence (5'-3')	Thermal Cycling Parameters
		***NECAP1* assays**	
NECAP1_F	Allelic discrimination assay	CAACCAAGAAGGGAGGTGTT	Listed in Table A5
NECAP1_R		GAGGATGGTGGGGGAATAGT	
NECAP1_Normal		HEX-CACAGGGGCT**G**GGGGTTTAA-IBFQ	
NECAP1_Alternate		FAM-AGACCAC**A**GGGGCTAGGGGTTT-IBFQ	
NECAP1_Finnish_seq_F	PCR and Sequencing	TGGTTTCTTTCCCCTACTCCC	Listed in Table A7
NECAP1_Finnish_seq__R		AATGGGTGGTGGAGTGACAT	
NECAP1_seq_F	PCR and Sequencing	GGCTTCAAGGAAGGACAGACT	Listed in Table A8
NECAP1_seq_R		GATACGATGATTCCTCAAAGTTAAA	
		Haplotype analysis SNV primers	
36894608_F	PCR and Sequencing	CAGGTGCCCTCAAAGAATCTA	Listed in Table A8
36894608_R		CAGAGCCTTGCTCTAACTCCA	
37020314_F	PCR and Sequencing	CTTGGAATGCTCCATCTCAAC	
37020314_R		CTGCATACTGGTTGGTCGATT	
37425357_F	PCR and Sequencing	TTCTTGATCTCCAGGGAGAGAG	
37425357_R		TCTTCCCTCTGGAAGAGGTTC	
37467780_F	PCR and Sequencing	TGGAACTTCCCATTCACTGTT	
37467780_R		GGTGCTATGCGTTACCACTGT	
37501948_F	PCR and Sequencing	CCTTCTAGACTCTTGGTAACCACTG	
37501948_R		GGGAATACTGGACAGTTACATGC	
37541149_F	PCR and Sequencing	AACGTAGGCCCAGTGACCTAT	
37541149_R		CCACAAGTGCTGGGAGATTTA	
37644083_F	PCR and Sequencing	AAGATCCAGTCTCTGGCTGGT	
37644083_R		CTCTCATTGCCACACAGTCAA	
37770618_F	PCR and Sequencing	GCCTATAAAGACAGTTCTCCCAAT	
37770618_R		GCACCTGCATTTGGTATCATT	

Table A3: Primers used in Chapter 4.

Primer Name	Assay Method Used	Primer Sequence (5'-3')	Thermal Cycling Parameters
IMPG2_Long_F	PCR and Gel	CCAGGCCTCATGTTTAATAGC	Listed in Table A9
IMPG2_Long_R		GCACTGTTGGGTTCTTGGATA	
IMPG2_LINE_Normal_F	PCR and AFLP	GATTTCAGGAAAGAAGTGAGAGG	Listed in Table A10
IMPG2_LINE_Alternate_F		TCTGAAGAAAGGAAAGGACACCAT	
IMPG2_LINE _R		FAM-ACCCAGAGAAATCTTCCCAAA	
IMPG2_intronic_SNV_F	Allelic discrimination assay	GGGACAAAGGAAACAGGTATCA	Listed in Table A11
IMPG2_ intronic_SNV_R		GAAGGAGAAGCAGGCTCCAT	
IMPG2_ intronic_SNV_Normal		HEX-CCCAGTGGCTCAGCGGTTTAG-IBFQ	
IMPG2_ intronic_SNV_Alternate		FAM-CCCAGTGGCTCAGCAGTTTAG-IBFQ	
GWAS_SNP_BICF2G630247609_F	PCR and Sequencing	TGAAGGTAAGTGTGTCATTTCTGA	Listed in Table A10
GWAS_SNP_BICF2G630247609_R		GTTAACCTTAACTACAATGACCAGTTT	
CEP97_intronic_SNV_F	PCR and Sequencing	CAATGAAATGTTAGCAGGGCTAC	Listed in Table A10
CEP97_intronic_SNV_R		GGAGAATCGGAGAATCCAGA	
IMPG2_cDNA_exons16-17_F	qPCR	GCCTGGTTTCTGCTTGAATG	Listed in Table A11
IMPG2_cDNA_exons16-17_R		TCACAATGCTTACCTCGGTAC	
IMPG2_cDNA_exons16-17_probe		FAM-TGTAGATGCCGGGTAGGTGAGAACT- IBFQ	
TBP_F	qPCR	AGCGAGGAAATATGCCAGAG	Listed in Table A12
TBP_R		GGGAACTTCACATCACAGCTC	
TBP_probe		FAM-TTCAAGATTCAGAACATGGTGGG- IBFQ	
IMPG2_promoter_F	PCR	tagtagGCTAGCGGATTTTCAAGAGAAATATGTTTTAGATC	Listed in Table A13
IMPG2_promoter_R		tagtagGGATCCTTGGGCCACAATCAAAGG	
Primers to verify sequence of the constructed pGL4 *IMPG2*-promoter plasmid			
RVprimer3_pp1_F		CTAGCAAAATAGGCTGTCCC	Directly Sanger sequenced (conditions and parameters listed in Chapter 2, section 2.5.)
pBR322ori-F		AAAGATACCAGGCGTTTCCC	

Table A3 (continued).

Primer Name	Assay Method Used	Primer Sequence (5'-3')	Thermal Cycling Parameters
L4440		CTCGCTCACTGACTCGCT	Directly Sanger sequenced (conditions and parameters listed in Chapter 2, section 2.5.)
SV40pA-R		GAAATTTGTGATGCTATTGC	
EBV Reverse		GTGGTTTGTCCAAACTCATC	
Seq_plas_1		GCTGAAGCCAGTTACCTTCG	
Seq_plas_2		GCGATCTGCCTATTTCGTTC	
Seq_plas_3		GAATGAGTGCGACACGAAAA	
Seq_plas_4		TACCTACGCCGAGTACTTCG	
Seq_plas_5		CACCTTCGTGACTTCCCATT	
Seq_plas_6		GAGCTATTCTTGCGCAGCTT	
Seq_plas_7		GAGCGGCTACGTTAACAACC	
Seq_plas_8		CGAGAAGGAGATCGTGGACT	
pp1_R		ACCACAATGGGACCAATAAA	
pp2_F		GGATTTTCAAGAGAAATATGTTTTAGATC	
pp2_R		TTAAACATGAGGCCTGGTATG	
pp3_F		TTTCTTGGTTAAGAGTTCTTCATGT	
pp3_R		ACCCTGAGATCACAATCTGG	
pp4_F		TATTCGGTCTGGGCTTACAG	
pp4_R		CAGCTATGAAAGCGGAAGAG	
pp5_F		TGCCAGTTAAAGGCAAAGAA	
pp5_R		TTGGGCCACAATCAAAGG	
pp6_F		TTTGGGAATGCTTGTCTGTT	
pp6_R		CGAAGTACTCGGCGTAGGTA	
pp7_F		CTAACTGGCCGGTACCTGAG	
pp7_R		TTTGGAATCTTGCTTCATTAGC	

Table A4: Primers used in Chapter 5.

Primer Name	Assay Method Used	Primer Sequence (5'-3')	Thermal Cycling Parameters
chr30:24115675_F	Multiplex PCR for NGS	CCAGAAATGTCAACTCCTTGG	Listed in Table A6
chr30:24115675_R		ATGAGCTGAAGCTCTGACTGC	
chr16:36561913_F		GCAACACATTGGGTGCATTAT	
chr16:36561913_R		GAAAGCAGGCTGAGGTACACA	
chr36:10135919_F		GCATGAGGAAGAAACTTGTCC	
chr36:10135919_R		GCTTAGGCTTGCATTTGTTTG	
chr39:43769675_F		TAAAGGACGCAGGGTTCTACA	
chr39:43769675_R		ACTGATGGAAGCTGACTTGGA	
chr1:110147994_F		ACTGATCCAGGCCCTGTTACT	
chr1:110147994_R		CTTTCTCCTTCCAGCCTCAAT	
chr39:1762952_F		ATCCGTCGATGTTGGAAGTG	
chr39:1762952_R		CAGTTGGAGGATGTGGGTGT	
chr2:62484625_F		GCCTGAGTTTGGAGAAGTCTG	
chr2:62484625_R		CTTGACCCAGAAGACCTCCTT	
chr17:40080783_F		GGGGCTGGAACTTCACTTTT	
chr17:40080783_R		CCTTACCTGGGGCTGGAG	
chr1:119547008_F		TTGATAGCGCAAAGGTGAAAC	
chr1:119547008_R		GGTTCCAACGACAATGAGAAA	
chr1:114617154_F		CCCAAATTTGACTGGGTGAG	
chr1:114617154_R		CCAGAGCCGACTTCGAGTG	
chr22:58226397_F		ACGGGCATGGACATGGAG	
chr22:58226397_R		CACCACCACCTGGTCAACCT	
chr4:68451901_F		AGACACAGGCAGAGGGAGAAG	
chr4:68451901_R		CCTCAGGCCCTGCTTATTC	
chr1:116038553_F		ACTCTAAGGCCAGAGCCACAC	
chr1:116038553_R		CGCCAACAGCTTACTTTTCA	

Table A4 (continued).

Primer Name	Assay Method Used	Primer Sequence (5'-3')	Thermal Cycling Parameters
chr16:44477295_F	Multiplex PCR for NGS (continued)	GGATGGCATCGGTACACATT	Listed in Table A6
chr16:44477295_R		ACATCTGAAAGCGGATTACCA	
chr15:52265347_F		TCAACCTGGATGAGACCTACC	
chr15:52265347_R		TTTGGACATCTGTCTCCCACT	
BBS2_F	Allelic Discrimination Assay	CCATCTTGGACTGGTTCTTG	Listed in Table A5
BBS2_R		CTGGAAAGGTTGTGAATGCT	
BBS2_Normal		HEX- ACACCATCATCCGAGCAGTACTGA - IBFQ	
BBS2_Alternate		FAM-ACACCATCATCCGACCAGTACTGA- IBFQ	

Table A5: Thermal cycling parameters for *NECAP1* and *BBS2* allelic discrimination assays.

PCR Stage	Temperature	Duration	Cycles
Pre-PCR step	25 °C	30 seconds	
Initial Denaturation	95 °C	3 minutes	
Denaturation	95 °C	3 seconds	35
Annealing/Extension	60 °C	10 seconds	
Post-PCR	25 °C	30 seconds	

Table A6: Thermal cycling parameters for multiplex PCR assays.

PCR Stage	Temperature (°C)	Duration	Cycles
Initial Denaturation	95	10 minutes	
Denaturation	95	15 seconds	25
Annealing	60	2 minutes	
Extension	60	30 minutes	

Table A7: Thermal cycling parameters for PCR amplification of the *NECAP1* variant in the Finnish cohort to perform Sanger sequencing.

PCR Stage	Temperature (°C)	Duration	Cycles
Initial Denaturation	95	5 minutes	
Denaturation	95	30 seconds	37
Annealing	59	30 seconds	
Extension	72	45 seconds	
Final Extension	72	10 minutes	

Table A8: Thermal cycling parameters for PCR amplification of the *NECAP1* variant in the AHT cohort to perform Sanger sequencing.

PCR Stage	Temperature (°C)	Duration	Cycles
Initial Denaturation	95	5 minutes	
Denaturation	95	30 seconds	35
Annealing	58	30 seconds	
Extension	72	30 seconds	
Final Extension	72	5 minutes	

Table A9: Thermal cycling parameters for PCR to characterise the *IMPG2* LINE-1 insertion.

PCR Stage	Temperature (°C)	Duration	Cycles
Denaturation	98	10 seconds	30
Annealing	60	15 seconds	
Final Extension	68	8 minutes	

Table A10: Thermal cycling parameters for *IMPG2* LINE-1 insertion, GWAS Top SNP (BICF2G630247609) and *CEP97* intronic SNV genotyping assays.

PCR Stage	Temperature (°C)	Duration	Cycles
Initial Denaturation	95	5 minutes	
Denaturation	95	30 seconds	35
Annealing	60	30 seconds	
Extension	72	30 seconds	
Final Extension	72	5 minutes	

Table A11: Thermal cycling parameters for *IMPG2* intronic SNV allelic discrimination assay.

PCR Stage	Temperature	Duration	Cycles
Pre-PCR step	25 °C	30 seconds	
Initial Denaturation	95 °C	3 minutes	
Denaturation	95 °C	3 seconds	40
Annealing/Extension	60 °C	10 seconds	
Post-PCR	25 °C	30 seconds	

Table A12: Thermal cycling parameters for qPCR assay to determine *IMPG2* expression.

PCR Stage	Temperature (°C)	Duration	Cycles
Initial Denaturation	95	5 minutes	
Denaturation	95	10 seconds	40
Annealing/Extension	60	30 seconds	

Table A13: Thermal cycling parameters for PCR amplification of the *IMPG2* promoter insert.

PCR Stage	Temperature (°C)	Duration	Cycles
Denaturation	98	10 seconds	30
Annealing	60	15 seconds	
Extension	68	8 minutes	

Appendix 5. In-house script used for *de novo* assembly of sequencing reads.

This script was written by Dr Oliver Forman for use within the AHT Canine Genetics group.

```
## Get reads containing substring AAA or its revcomp TTT

use strict;

use Getopt::Std ;

use File::Basename ;

####################

# Define variables #

####################

my $command  = "";

my $in1             = "";

my $in2             = "";

my $fastq    = "";

my $out             = "";

system clear;

use Term::ANSIColor;

print color "bold cyan";

print " \n\n\n\n\n  FASTQ FILTER!\n\n\n";

print color "bold white";

print "  THIS IS A SHORT SCRIPT TO OUTPUT A SUBSET OF READS CONTAINING A PARTICULAR
SEQUENCE\n";

print color 'reset';

print "  Oliver Forman Aug2014\n\n\n\n\n\n";

print "\n\nEnter a short name for this analysis:\n\n    ";

$out = <STDIN>;

chomp $out;

print "\n\nEnter the names of your fastq file including .fastq extention:\n\n    ";

$fastq = <STDIN>;

chomp $fastq;
```

Appendix 5 (continued).

```
print "\n\nPlease enter the nucleotide sequence you want the reads to contain:\n\n
";

$in1 = <STDIN>;

chomp $in1;

print "\n\nPlease enter the reverse complement of the nucleotide sequence:\n\n    ";

$in2 = <STDIN>;

chomp $in2;

print "\n\n\n\nThe program is running, run time is dependant on the FASTQ file
size....";

$command =  "paste - - - - < $fastq|grep -P '^@.*?\t(.*?$in1.*?)|(.*?$in2.*?)\t\+'|tr
'\t' '\n'print > $out.fastq ";

system("$command");

print "FINISHED!\n\n";
```

Appendix 6. In-house effect scores assigned to sequence ontology terms used to prioritise variant filtration in whole genome sequencing analysis.

Sequence ontology term	Effect Score
5'-UTR premature start codon gain variant	5
Chromosome number variation	5
Disruptive in-frame deletion	5
Disruptive in-frame insertion	5
Exon loss	5
Exon loss variant	5
Frameshift variant	5
In-frame deletion	5
In-frame insertion	5
Initiator codon variant	5
Initiator codon variant and non-canonical start codon	5
Missense variant	5
Protein altering variant	5
Rare amino acid variant	5
Splice acceptor variant	5
Splice donor variant	5
Start lost	5
Stop gained	5
Stop lost	5
Transcript ablation	5
3'-UTR truncation	4
5 '-UTR truncation	4
Non-coding exon variant	4
Regulatory region ablation	4
Splice region variant	4
TFBS ablation	4
Coding sequence variant	3
Conserved intergenic variant	3
Conserved intron variant	3
Non-coding RNA transcript variant	3
3 prime UTR variant	2
5 prime UTR variant	2
Exon variant	2
Feature elongation	2
Feature truncation	2
Gene variant	2
Intron variant	2
Mature microRNA (miRNA) variant	2
miRNA	2
Nonsense mediated decay (NMD) transcript variant	2
Non-coding transcript exon variant	2
Non-coding transcript variant	2
Regulatory region amplification	2
Regulatory region variant	2
Start retained	2
Stop retained variant	2
TFBS variant	2
TFBS amplification	2
Transcript amplification	2
Transcript variant	2
Upstream gene variant	2
Coding sequence variant	1
Downstream gene variant	1
Intergenic region	1
Intergenic variant	1
Intragenic variant	1
Synonymous variant	1
Transcript	1

Appendix 7. Breed names known for canids screened for the *NECAP1* variant.

Breed	Number of individuals genotyped
Affenpinscher	5
Afghan Hound	3
Airedale Terrier	5
Alaskan Malamute	6
Alpine Dachsbracke	1
American Bulldog	5
American Cocker Spaniel	3
American Staffordshire Terrier	1
Australian Cattle Dog	6
Australian Kelpie	3
Australian Shepherd	3
Australian Terrier	1
Basset Fauve De Bretagne	3
Basset Hound	8
Bavarian Hound	1
Beagle	10
Bearded Collie	16
Bedlington Terrier	1
Belgian Shepherd Dog (Groenendael)	1
Belgian Shepherd Dog (Tervueren)	2
Berger Blanc Suisse	1
Berger Picard (Picardy sheepdog)	4
Bichon Frisé	4
Bloodhound	4
Border Collie	44
Border Terrier	8
Borzoi	3
Boston Terrier	3
Boxer	5
Briard	2
Brittany Spaniel	3
Brussels Griffon	1
Bull Dog	6
Bull Terrier	6
Bullmastiff	2
Cairn Terrier	2
Cane Corso	1
Catalan Sheepdog	3
Cavalier King Charles Spaniel	7
Central Asian Shepherd dog	1
Cesky Terrier	3
Chesapeake Bay Retriever	4
Chihuahua	2
Chinese Crested	1
Chinese Indigenous dogs	28
Chow Chow	5
Cocker Spaniel	4
Curly Coated Retriever	4
Dachshund	3
Dalmatian	7
Dandie Dinmont Terrier	7
Doberman	7
Dogue de Bordeaux	4
Dutch Shepherd	1
Elo	1
English Setter	4
English Springer Spaniel	10
Entlebucher Sennenhund	8
Eurasier	2
Field Spaniel	4
Finnish Lapphund	5

Appendix 7 (continued).

Flat-Coated Retriever	6
French Bull Dog	10
German Pinscher	3
German Shepherd	16
German Spitz (Giant)	116
German Spitz (Medium)	151
German Spitz (Miniature)	117
German Spitz (Pomeranian)	59
German Wire-Haired Pointer	4
Giant Schnauzer	323
Glen of Imaal Terrier	3
Golden Retriever	9
Gordon Setter	3
Grand Basset Griffon Vendéen	3
Great Dane	4
Great Pyrenees	1
Greater Swiss Mountain Dog	7
Greyhound	5
Havanese	3
Heideterrier	1
Hovawart	4
Hungarian Vizsla	7
Hungarian Wire-Haired Vizsla	3
Icelandic Sheepdog	3
Irish Red and White Setter	13
Irish Setter	3
Irish Soft-Coated Wheaten Terrier	1
Irish Terrier	3
Irish Water Spaniel	3
Irish Wolfhound	8
Italian Greyhound	4
Italian Spinone	5
Jack Russell Terrier	4
Jagdterrier	1
Japanese Akita	4
Japanese Chin	1
Japanese Shiba Inu	3
Karelian Bear Dog	1
Keeshond	3
Kerry Blue Terrier	1
Kromfohrländer	1
Kunming Dog	10
Labrador Retriever	18
Lagotto Romagnolo	9
Lakeland Terrier	4
Lancashire Heeler	3
Landseer	2
Large Munsterlander	4
Leonberger	50
Lhasa Apso	9
Malinois	6
Manchester Terrier	3
Miniature Smooth-Haired Dachshund	3
Miniature Bull Terrier	7
Miniature Bullterrier	1
Miniature Long-Haired Dachshund	168
Miniature Poodle	2
Miniature Schnauzer	89
Miniature Wire-Haired Dachshund	3
Newfoundland	1
Norfolk Terrier	3
Northern Inuit	5
Norwegian Buhund	7
Norwegian Elkhound	3
Norwich Terrier	7
Nova Scotia Duck Tolling Retriever	4

Appendix 7 (continued).

Old English Mastiff	3
Old English Sheepdog	5
Otterhound	3
Parson Russel Terrier	4
Pembroke Welsh Corgi	3
Petit Basset Griffon Vendéen	6
Pharaoh Hound	3
Polish Lowland Sheepdog	3
Poodle	19
Portuguese Water Dog	3
Pug	21
Rhodesian Ridgeback	7
Rottweiler	2
Rough Collie	3
Saint Bernard	3
Saluki	5
Samoyed	3
Scottish Deerhound	3
Scottish Terrier	7
Shar Pei	3
Shetland Sheepdog	5
Shih Tzu	11
Siberian Husky	8
Skye Terrier	5
Sloughi	3
Soft-Coated Wheaten Terrier	3
Spanish Water Dog	1
St. Bernard	2
Staffordshire Bull Terrier	4
Standard Long-Haired Dachshund	3
Standard Poodle	4
Standard Schnauzer	3
Standard Smooth-Haired Dachshund	4
Standard Wire-Haired Dachshund	3
Swedish Vallhund	6
Tibetan Mastiff	10
Tibetan Spaniel	10
Tibetan Terrier	5
Utonagan	3
Weimaraner	5
Welsh Springer Spaniel	9
West Highland White Terrier	20
Whippet	5
White Shepherd	1
Working Sheepdog	3
Yorkshire Terrier	67
Cross breeds	10
Wolf	3
Unknown/unconfirmed breeds	3,156
Total	**5,130**

Appendix 8. *IMPG2* LINE-1 insertion sequence.

GAAATCTTCCCAAAATGAGGAAACATAATCATTTGGGCCACAATCAAAGGAATGAGGAGAAGCCAAAA
TCCTTAACTTGAACAACCCAAAAGGCTTGAAACTTCCAGTAACAAATTATTATAGGAAAGGTCTAACA
TTTAAGAGTTTATGTTTGAGAACAGACAAGCATTCCCAAACCATTATAAATTAGCTTAAATAAAAATG
GTGTCCTTTCCTTTCTTCAGAGAGTTAGAACAAATTATTTTAAGATTTGTGTGGAATCAGAAAAGACC
CCGAATAGCCAGGGGAATTTTAAAAAAGAAAACCATATCTGGGGGCATCACAATGCCAGATTTCAGGT
TGTACTACAAAGCTGTGGTCATCAAGACAGTGTGGTATTGGCACAAAAACAGACACATAGATCAGTGG
AACAGAATAGAGAATCCAGAAGTGGACCCTGAACTTTATGGGCAACTAATATTCGATAAAGGAGGAAA
GACTATCCATTGGAAGAAAGACAGTCTCTTCAATAAATGGTGCTGGGAAAATTGGACATCCACATGCA
GAAGAATGAAACTAGACCACTCTCTTTCACCATACACAAAGATAAACTCAAAATGGATGAAAGATCTA
AATGTGAGACAAGATTCCATCAAAATCCTAGAGAAGAACACAGGCAACACCCTTTTTGAACTCGGCCA
TAGTAACTTCTTGCAAGATACATCCACGAAGGCAAAAGAAACAAAAGCAAAAATGAACTATTGGGACT
TCATCAAGATAAGAAGCTTTTGCACAGCAAAGGATACAGTCAACAAAACTCAAAAACAACCTACAGAA
TGGGAGAAGATATTTGCAAATGACATATCAGATAAAGGGCTAGTTTCCAAGATCTATAAAGAACTTAT
TAAACTCAACACCAAAGAAACAAACAATCCAATCATGAAATGGGCAAAAGACATGAACAGAAATCTCA
CAGAGGAAGACATAGACATGGCCAACATGCATATGAGAAAATGCTCTGCATCACTTGCCATCAGGGAA
ATACAAATGAAAACTACAATGAGATACCACCTCACACCAGTGAGAATGGGGAAAATTATCAAGGCAGG
AAACAACAAATGTTGGAGAGGATGCGGAGAAAAGGGAACCCTCTTACACTGTTGGTGGGAATGTGAAC
TGGTGCAGCCACTCTGGAAAACTGTGTGGAGGTTCCTCAAACAGTTAAAAATATACCTGCCCTACGAC
CCAGCAATTGCACTGTTGGGGATTTACCCCAAAGATACAAATGCAATGAAACGCCGGGACACCTGCAC
CCCGATGTTTCTAGCAGCAATGGCCACGATAGCCAAACTGTGGAAGGAGCCTCGGTGTCCAACGAAAG
ATGAATGGATAAAGAAGATGTGGTTTATGTATACAATGGAATATTACTCAGCTATTAGAAATGACAA
ATACCCACCATTTGCTTCAACGTGGATGGAACTGGAGGGTATTATGCTGAGTGAAGTAAGTCAGTCAG
AGAAGGACAAACATTATATGTTCTCATTCATTTGGGGAATATAAATAATAGTGAAAGGGAAAATAAG
GGAAGGGAGAAGAAATGTGTGGGAAATATCAGAAAGGGAGACAGAACATAAAGACTGCTAACTCTGG
GAAACGAACTAGGGGTGGTAGAAGGGGAGGAGGGCGGGGCGTGGGAGTGAATGGGTGACGGGCACTG
GGTGTTATTCTGTATGTTAGTAAATTGAACACCAATAAAAAAATAAATAAAAATAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGGTGTCCTTTCAAAATACC
TCTCACTTCTTTC

Flanking *IMPG2* sequence

16 base pair target site duplication

LINE-1 insertion sequence

Poly-A tail

Flanking *IMPG2* sequence

Appendix 9. Breed names for dogs screened for the *IMPG2* LINE-1 insertion.

Breed	Number of individuals genotyped
American Bull Dog	3
Australian Cattle Dog	3
American Cocker Spaniel	3
Affenpinscher	3
Afghan Hound	3
Australian Kelpie	3
Alaskan Malamute	3
Australian Shepherd	3
Airedale Terrier	3
Boxer	3
Basset Hound	3
Border Collie	3
Bulldog	3
Beagle	3
Bearded Collie	3
Basset Fauve De Bretagne	3
Bloodhound	3
Border Terrier	3
Borzoi	3
Berger Picard (Picardy sheepdog)	3
Brittany Spaniel	3
Belgian Shepherd Dog (Tervueren)	2
Belgian Shepherd Dog (Groenendael)	1
Boston Terrier	3
Bull Terrier	3
Chesapeake Bay Retriever	3
Curly Coated Retriever	3
Cesky Terrier	3
Chow Chow	3
Cavalier King Charles Spaniel	3
Cocker Spaniel	3
Catalan Sheepdog	3
Doberman	3
Dandie Dinmont	3
Dalmatian	3
English Setter	3
English Springer Spaniel	3
French Bull Dog	3
Flat-Coated Retriever	3
Finnish Lapphund	3
Field Spaniel	3
Greyhound	3
Grand Basset Griffon Vendéen	3
Great Dane	3
Glen of Imaal Terrier	3
Gordon Setter	3
German Pinscher	3
Golden Retriever	3
Giant Schnauzer	3
German Shepherd Dog	3

Appendix 9 (continued).

Greater Swiss Mountain Dog	3
German Spitz (Klein)	3
German Wire-Haired Pointer	3
Havanese	3
Hovawart	3
Icelandic Sheepdog	3
Italian Greyhound	3
Irish Red and White Setter	3
Irish Setter	3
Italian Spinone	3
Irish Water Spaniel	3
Japanese Akita	3
Jack Russell Terrier	3
Japanese Shiba Inu	3
Keeshond	3
Lhasa Apso	63
Leonberger	3
Lancashire Heeler	3
Large Münsterländer	3
Labrador Retriever	3
Lagotto Romagnolo	2
Lakeland Terrier	3
Miniature Bull Terrier	3
Miniature Long-Haired Dachshund	3
Miniature Schnauzer	3
Miniature Smooth-Haired Dachshund	3
Manchester Terrier	1
Miniature Wire-Haired Dachshund	**3**
Norwegian Buhund	3
Norwegian Elkhound	3
Northern Inuit	3
Norfolk Terrier	3
Nova Scotia Duck Tolling Retriever	3
Norwich Terrier	3
Old English Mastiff	3
Old English Sheepdog	3
Otterhound	3
Pharaoh Hound	3
Polish Lowland Sheepdog	3
Parson Russell Terrier	3
Petit Basset Griffon Vendéen	3
Pug	2
Pembroke Welsh Corgi	3
Rough Collie	3
Rhodesian Ridgeback	3
Standard Schnauzer	3
Saluki	3
Samoyed	3
Staffordshire Bull Terrier	3
Scottish Deerhound	3
Scottish Terrier	3
Soft-coated Wheaten Terrier	3
Shar Pei	3
Siberian Husky	2

Appendix 9 (continued).

Standard Long-Haired Dachshund	3
Standard Smooth-Haired Dachshund	3
Standard Poodle	3
Shetland Sheepdog	3
Skye Terrier	3
Saint Bernard	3
Shih Tzu	20
Swedish Vallhund	3
Tibetan Spaniel	10
Tibetan Terrier	5
Utonagan	3
Vizsla (Smooth-Haired)	3
Vizsla (Wire-Haired)	3
Weimaraner	3
Wire-haired Dachshund	3
West Highland White Terrier	3
Whippet	3
Working Sheepdog	3
Welsh Springer Spaniel	3
Total	**447**

Appendix 10. Transcription factor binding site predictions known to be essential for normal eye development or function and are situated within 150-bp upstream and 150-bp downstream of the *IMPG2* LINE-1 breakpoints.

Five of these transcription factor binding site predictions are specific to photoreceptor cells in the retina (highlighted in orange).

MatInspector Matrix Family	Detailed Family Information	Detailed Matrix Information	Sequence
V$HBOX	Homeobox transcription factors	Homeobox protein engrailed (en-1)	tctttgccTTTAactggca
V$BCDF	Bicoid-like homeodomain transcription factors	Pituitary Homeobox 1 (Ptx1, Pitx-1)	ccagCTAAgctctttat
V$BCDF	Bicoid-like homeodomain transcription factors	Cone-rod homeobox-containing transcription factor / otx-like homeobox gene	tattcTAATctagccag
V$CART	Cart-1 (cartilage homeoprotein 1)	Binding site for S8 type homeodomains	taagcTAATtagaacagatgg
V$BCDF	Bicoid-like homeodomain transcription factors	Photoreceptor conserved element 1	taagcTAATtagaacag
V$BRNF	Brn POU domain factors	POU class 3 homeobox 2 (POU3F2), OTF7	tctgttcTAATtagcttat
V$HBOX	Homeobox transcription factors	Homeodomain transcription factor Gsh-2	ataagcTAATtagaacaga
V$LHXF	Lim homeodomain factors	LIM homeobox 8	catctgttctAATTagcttatct
V$PAXH	PAX homeodomain binding sites	Paired box 4, homeodomain binding site	aagctAATTagaaca
V$BCDF	Bicoid-like homeodomain transcription factors	Photoreceptor conserved element 1	tgttcTAATtagcttat
V$BRNF	Brn POU domain factors	POU class 3 homeobox 2 (POU3F2), OTF7	gataagcTAATtagaacag
V$HBOX	Homeobox transcription factors	Homeodomain transcription factor Gsh-2	ctgttcTAATtagcttatc
V$LHXF	Lim homeodomain factors	LIM homeobox 6	gagataagcTAATtagaacagat
V$PAXH	PAX homeodomain binding sites	Paired box 4, homeodomain binding site	gttctAATTagctta
V$CART	Cart-1 (cartilage homeoprotein 1)	Binding site for S8 type homeodomains	tgttcTAATtagcttatctct
V$LHXF	Lim homeodomain factors	LIM-homeodomain transcription factor LHX3	ggaagagataagcTAATtagaac
V$PAX6	PAX-4/PAX-6 paired domain binding sites	PAX6 paired domain and homeodomain are required for binding to this site	tcctgaaatCCAGcacagc
V$FKHD	Fork head domain factors	Fork head homologous X binds DNA with a dual sequence specificity (FHXA and FHXB)	cctttcAAAAtacctct
V$NR2F	Nuclear receptor subfamily 2 factors	Chicken ovalbumin upstream promoter transcription factor 2, NR2F2 homodimer, DR1 sites	gaggtattttgaaaGGACaccattt
V$CART	Cart-1 (cartilage homeoprotein 1)	Prophet of Pit 1, PROP paired-like homeobox 1	tagcttAAATaaaaatggtgt
V$LHXF	Lim homeodomain factors	LIM homeobox transcription factor 1, beta	caccatttttATTTaagctaatt
V$LHXF	Lim homeodomain factors	LIM homeobox transcription factor 1, alpha	aaattagcttAAATaaaaatggt
V$LHXF	Lim homeodomain factors	LIM-homeodomain transcription factor LHX3	tttttatttaagcTAATttataa
V$CART	Cart-1 (cartilage homeoprotein 1)	Aristaless-like homeobox 3	attataAATTagcttaaataa
V$HBOX	Homeobox transcription factors	Engrailed homeobox 2	cattataAATTagcttaaa
V$LHXF	Lim homeodomain factors	LIM homeobox 6	tatttaagcTAATttataatggt
V$PAXH	PAX homeodomain binding sites	Paired box 6, homeodomain binding site	ttataAATTagctta
V$BCDF	Bicoid-like homeodomain transcription factors	Photoreceptor conserved element 1	taagcTAATttataatg
V$BRNF	Brn POU domain factors	POU class 4 homeobox 3 (POU4F3), BRN3C	ccattatAAATtagcttaa
V$HBOX	Homeobox transcription factors	Homeodomain transcription factor Gsh-2	ttaagcTAATttataatgg
V$LHXF	Lim homeodomain factors	LIM homeobox 1	aaccattataAATTagcttaaat
V$CART	Cart-1 (cartilage homeoprotein 1)	Phox2a (ARIX) and Phox2b	taagcTAATttataatggttt
V$BRNF	Brn POU domain factors	Brn-3, POU-IV protein class	agctaatttATAAtggttt
V$FKHD	Fork head domain factors	HNF-3/Fkh Homolog 1 (FOXQ1)	caaacaTAAActcttaa
V$PAX6	PAX-4/PAX-6 paired domain binding sites	Pax-6 paired domain binding site	cataaACTCttaaatgtta
V$FKHD	Fork head domain factors	Fork head homologous X binds DNA with a dual sequence specificity (FHXA and FHXB)	tttcctATAAtaatttg
V$BRNF	Brn POU domain factors	Brn-3, POU-IV protein class	tttcctaTAATaatttgtt
V$BRNF	Brn POU domain factors	Brn-3, POU-IV protein class	cctataaTAATttgttact
V$HBOX	Homeobox transcription factors	Engrailed homeobox 1	agtaacaAATTattatagg
V$PAXH	PAX homeodomain binding sites	Paired box 6, homeodomain binding site	taacaAATTattata
V$BRNF	Brn POU domain factors	Brn-3, POU-IV protein class	cagtaacaaATTAttatag
V$CART	Cart-1 (cartilage homeoprotein 1)	HESX homeobox 1, dimeric binding site	tataaTAATttgttactggaa

Appendix 11. Breed names for canids screened for the *BBS2* variant.

Breed	Number of individuals genotyped
Affenpinscher	2
Airedale Terrier	5
Alaskan Malamute	4
Alpine Dachsbracke	2
American Bull Dog	2
American Cocker Spaniel	1
American Staffordshire Terrier	2
Australian Cattle Dog	5
Australian Shepherd	3
Australian Terrier	1
Basenji	7
Basset Hound	9
Bavarian Hound	1
Beagle	8
Bearded Collie	13
Bedlington Terrier	1
Berger Blanc Suisse	1
Berger Picard	1
Bichon Frisé	6
Black Russian Terrier	1
Bloodhound	1
Border Collie	136
Border Terrier	10
Boxer	2
Briard	2
Brussels Griffon	3
Bull Terrier	3
Bulldog	3
Bullmastiff	2
Cairn Terrier	4
Cane Corso	5
Cardigan Welsh Corgi	1
Cavalier King Charles Spaniel	6
Central Asian Shepherd Dog	1
Cesky Terrier	1
Chesapeake Bay Retriever	1
Chihuahua	6
Chinese Crested	3
Chinese Indigenous Dog	28
Chow Chow	2
Cocker Spaniel	1
Collie	1
Corgi	1
Curly Coated Retriever	5
Dachshund	4
Dalmatian	4
Dandie Dinmont Terrier	5
Doberman Pinscher	5
Dogue de Bordeaux	6
Dutch Shepherd	1

Appendix 11 (continued).

Elo	1	Heideterrier	1
English Bull Dog	1	Hovawart	3
English Cocker Spaniel	3	Hungarian Vizsla	4
English Mastiff	2	Hungarian Wire-Haired Vizsla	1
English Setter	1	Irish Red and White Setter	4
English Springer Spaniel	3	Irish Setter	2
Entlebucher Sennenhund	8	Irish Soft Coated Wheaten Terrier	1
Eurasier	2	Irish Terrier	3
Field Spaniel	1	Irish Water Spaniel	1
Finnish Lapphund	2	Irish Wolfhound	8
Flat Coated Retriever	4	Italian Greyhound	1
French Bull Dog	7	Italian Spinone	2
Friesian Stabyhoun	2	Jack Russell Terrier	8
German Shepherd	20	Jagdterrier	2
German Wirehaired	1	Japanese Akita	1
Giant Schnauzer	3	Japanese Chin	1
Giant Spitz	1	Karelian Bear Dog	1
Glen of Imaal Terrier	1	Keeshond	3
Golden Retriever	13	Kerry Blue Terrier	2
Grand Basset Griffon Vendéen	1	Kromfohrländer	1
Great Dane	2	Kunming Dog	10
Great Pyrenees	1	Labrador Retriever	14
Greater Swiss Mountain Dog	6	Lagotto Romagnolo	10
Greyhound	7	Lakeland Terrier	1
Griffon Bruxellois	1	Lancashire Heeler	2
Havanese	3	Landseer	2

Appendix 11 (continued).

Large Münsterländer	1
Leonberger	55
Lhasa Apso	1
Malinois	7
Maltese	1
Miniature Bull Terrier	2
Miniature Long- Haired Dachshund	3
Miniature Poodle	2
Miniature Schnauzer	25
Miniature Wire-Haired Dachshund	1
Newfoundland	3
Northern Inuit	2
Norwegian Buhund	3
Norwich Terrier	5
Nova Scotia Duck Tolling Retriever	2
Old English Sheepdog	2
Otterhound	2
Papillon	3
Pembroke Welsh Corgi	3
Petit Basset Griffon Vendéen	5
Pomeranian	4
Poodle	19
Portuguese Podengo	1
Portuguese Water Dog	3
Pug	22
Rhodesian Ridgeback	4

Rottweiler	4
Rough Collie	1
Saluki	2
Scottish Deerhound	3
Scottish Terrier	6
Shar Pei	2
Shetland Sheepdog	505
Shih Tzu	4
Siberian Husky	5
Skye Terrier	2
Sloughi	3
Smooth Collie	1
Soft-Coated Wheaten Terrier	2
Spanish Water Dog	1
St. Bernard	2
Staffordshire Bull Terrier	1
Standard Poodle	2
Standard Schnauzer	1
Swedish Vallhund	3
Tibetan Mastiff	10
Tibetan Terrier	3
Weimaraner	3
Welsh Springer Spaniel	9
West Highland White Terrier	20
Whippet	3
White Shepherd	1

Appendix 11 (continued).

Yorkshire Terrier	69
Cross Breed	15
Wolf	8
Total	**1,386**

Bibliography

Acharya, S., Foletta, V. C., Lee, J. W., Rayborn, M. E., Rodriguez, I. R., Young, W. S., 3rd & Hollyfield, J. G. (2000). SPACRCAN, a novel human interphotoreceptor matrix hyaluronan-binding proteoglycan synthesized by photoreceptors and pinealocytes. *The Journal of Biological Chemistry*, 275(10), 6945-55.

Acland, G. M. & Aguirre, G. D. (1987). Retinal degenerations in the dog: IV. Early retinal degeneration (erd) in Norwegian elkhounds. *Experimental Eye Research*, 44(4), 491-521.

Acland G. M. (1988). Diagnosis and differentiation of retinal diseases in small animals by electroretinography. *Seminars in veterinary medicine and surgery (small animal)*, 3(1), 15–27.

Acland, G. M., Aguirre, G. D., Ray, J., Zhang, Q., Aleman, T. S., Cideciyan, A. V., Pearce-Kelling, S. E., Anand, V., Zeng, Y., Maguire, A. M., Jacobson, S. G., Hauswirth, W. W. & Bennett, J. (2001). Gene therapy restores vision in a canine model of childhood blindness. *Nature Genetics*, 28(1), 92-5.

Acland, G. M., Aguirre, G. D., Bennett, J., Aleman, T. S., Cideciyan, A. V., Bennicelli, J., Dejneka, N. S., Pearce-Kelling, S. E., Maguire, A. M., Palczewski, K., Hauswirth, W. W. & Jacobson, S. G. (2005). Long-term restoration of rod and cone vision by single dose rAAV-mediated gene transfer to the retina in a canine model of childhood blindness. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, 12(6), 1072-82.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), 248-9.

Aguirre, G. (1978). Retinal degenerations in the dog. I. Rod dysplasia. *Experimental Eye Research*, 26(3), 233-53.

Aguirre, G., Alligood, J., O'Brien, P. & Buyukmihci, N. (1982). Pathogenesis of progressive rod-cone degeneration in miniature poodles. *Investigative Ophthalmology & Visual Science*, 23(5), 610-30.

Aguirre, G. D., Baldwin, V., Pearce-Kelling, S., Narfström, K., Ray, K. & Acland, G. M. (1998). Congenital stationary night blindness in the dog: common mutation in the RPE65 gene indicates founder effect. *Molecular Vision*, 4, 23.

Ahonen, S. J., Arumilli, M. & Lohi, H. (2013). A CNGB1 frameshift mutation in Papillon and Phalene dogs with progressive retinal atrophy. *PLoS One*, 8(8), e72122.

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., Garcia Giron, C., Hourlier, T., Howe, K., Kahari, A., Kokocinski, F., Martin, F. J., Murphy, D. N., Nag, R., Ruffier, M., Schuster, M., Tang, Y. A., Vogel, J. H., White, S., Zadissa, A., Flicek, P. & Searle, S. M. (2016). The Ensembl gene annotation system. *Database (Oxford)*, 2016.

Alazami, A. M., Hijazi, H., Kentab, A. Y. & Alkuraya, F. S. (2014). NECAP1 loss of function leads to a severe infantile epileptic encephalopathy. *Journal of Medical Genetics*, 51(4), 224-8.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. & Walter, P. (2014). *Molecular Biology of the Cell, Sixth Edition*Taylor & Francis Group.

Aldahmesh, M. A., Li, Y., Alhashem, A., Anazi, S., Alkuraya, H., Hashem, M., Awaji, A. A., Sogaty, S., Alkharashi, A., Alzahrani, S., Al Hazzaa, S. A., Xiong, Y., Kong, S., Sun, Z. & Alkuraya, F. S. (2014). IFT27, encoding a small GTPase component of IFT particles, is mutated in a consanguineous family with Bardet-Biedl syndrome. *Human Molecular Genetics,* 23(12), 3307-3315.

Alsahli, S., Al-Twaijri, W. & Al Mutairi, F. (2018). Confirming the pathogenicity of NECAP1 in early onset epileptic encephalopathy. *Epilepsia Open*, 3(4), 524-527.

American Kennel Club (2007). *The Complete Dog Book: 20th Edition,* Random House Publishing Group.

Ansley, S. J., Badano, J. L., Blacque, O. E., Hill, J., Hoskins, B. E., Leitch, C. C., Chul Kim, J., Ross, A. J., Eichers, E. R., Teslovich, T. M., Mah, A. K., Johnsen, R. C., Cavender, J. C., Alan Lewis, R., Leroux, M. R., Beales, P. L. & Katsanis, N. (2003). Basal body dysfunction is a likely cause of pleiotropic Bardet–Biedl syndrome. *Nature*, 425(6958), 628-633.

Applied Biosystems, T. F. S. (2020). *Real-time PCR: understanding C_t*. Available online: https://www.thermofisher.com/content/dam/LifeTech/Documents/PDFs/PG1503-PJ9169-CO019879-Re-brand-Real-Time-PCR-Understanding-Ct-Value-Americas-FHR.pdf (Accessed: 27 February 2020).

Arno, G., Agrawal, S. A., Eblimit, A., Bellingham, J., Xu, M., Wang, F., Chakarova, C., Parfitt, D. A., Lane, A., Burgoyne, T., Hull, S., Carss, K. J., Fiorentino, A., Hayes, M. J., Munro, P. M., Nicols, R., Pontikos, N., Holder, G. E., Ukirdc, Asomugha, C., Raymond, F. L., Moore, A. T., Plagnol, V., Michaelides, M., Hardcastle, A. J., Li, Y., Cukras, C., Webster, A. R., Cheetham, M. E. & Chen, R. (2016). Mutations in REEP6 Cause Autosomal-Recessive Retinitis Pigmentosa. *American Journal of Human Genetics*, 99(6), 1305-1315.

Artero Castro, A., Long, K., Bassett, A., Machuca, C., Leon, M., Avila-Fernandez, A., Corton, M., Vidal-Puig, T., Ayuso, C., Lukovic, D. & Erceg, S. (2019). Generation of gene-corrected human induced

pluripotent stem cell lines derived from retinitis pigmentosa patient with Ser331Cysfs*5 mutation in MERTK. *Stem Cell Research*, 34, 101341.

Baala, L., Audollent, S., Martinovic, J., Ozilou, C., Babron, M. C., Sivanandamoorthy, S., Saunier, S., Salomon, R., Gonzales, M., Rattenberry, E., Esculpavit, C., Toutain, A., Moraine, C., Parent, P., Marcorelles, P., Dauge, M. C., Roume, J., Le Merrer, M., Meiner, V., Meir, K., Menez, F., Beaufrere, A. M., Francannet, C., Tantau, J., Sinico, M., Dumez, Y., MacDonald, F., Munnich, A., Lyonnet, S., Gubler, M. C., Genin, E., Johnson, C. A., Vekemans, M., Encha-Razavi, F. & Attie-Bitach, T. (2007). Pleiotropic effects of CEP290 (NPHP6) mutations extend to Meckel syndrome. *American Journal of Human Genetics*, 81(1), 170-9.

Badano, J. L., Ansley, S. J., Leitch, C. C., Lewis, R. A., Lupski, J. R. & Katsanis, N. (2003). Identification of a Novel Bardet-Biedl Syndrome Protein, BBS7, That Shares Structural Features with BBS1 and BBS2. *American Journal of Human Genetics,* 72(3), 650-658.

Bainbridge, J. W., Smith, A. J., Barker, S. S., Robbie, S., Henderson, R., Balaggan, K., Viswanathan, A., Holder, G. E., Stockman, A., Tyler, N., Petersen-Jones, S., Bhattacharya, S. S., Thrasher, A. J., Fitzke, F. W., Carter, B. J., Rubin, G. S., Moore, A. T. & Ali, R. R. (2008). Effect of gene therapy on visual function in Leber's congenital amaurosis. *The New England Journal of Medicine*, 358(21), 2231-9.

Bandah-Rozenfeld, D., Collin, R. W., Banin, E., van den Born, L. I., Coene, K. L., Siemiatkowska, A. M., Zelinger, L., Khan, M. I., Lefeber, D. J., Erdinest, I., Testa, F., Simonelli, F., Voesenek, K., Blokland, E. A., Strom, T. M., Klaver, C. C., Qamar, R., Banfi, S., Cremers, F. P., Sharon, D. & den Hollander, A. I. (2010). Mutations in IMPG2, encoding interphotoreceptor matrix proteoglycan 2, cause autosomal-recessive retinitis pigmentosa. *American Journal of Human Genetics*, 87(2), 199-208.

Bax, N. M., Sangermano, R., Roosing, S., Thiadens, A. A., Hoefsloot, L. H., van den Born, L. I., Phan, M., Klevering, B. J., Westeneng-van Haaften, C., Braun, T. A., Zonneveld-Vrieling, M. N., de Wijs, I., Mutlu, M., Stone, E. M., den Hollander, A. I., Klaver, C. C., Hoyng, C. B. & Cremers, F. P. (2015). Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. *Human Mutation*, 36(1), 43-7.

Beales, P. L., Warner, A. M., Hitman, G. A., Thakker, R. & Flinter, F. A. (1997). Bardet-Biedl syndrome: a molecular and phenotypic study of 18 families. *Journal of Medical Genetics,* 34(2), 92-98.

Beales, P. L., Elcioglu, N., Woolf, A. S., Parker, D. & Flinter, F. A. (1999). New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *Journal of Medical Genetics*, 36(6), 437-446.

Bech-Hansen, N. T., Naylor, M. J., Maybaum, T. A., Pearce, W. G., Koop, B., Fishman, G. A., Mets, M., Musarella, M. A. & Boycott, K. M. (1998). Loss-of-function mutations in a calcium-channel alpha1-subunit gene in Xp11.23 cause incomplete X-linked congenital stationary night blindness. *Nature Genetics*, 19(3), 264-7.

Bech-Hansen, N. T., Naylor, M. J., Maybaum, T. A., Sparkes, R. L., Koop, B., Birch, D. G., Bergen, A. A., Prinsen, C. F., Polomeno, R. C., Gal, A., Drack, A. V., Musarella, M. A., Jacobson, S. G., Young, R. S. & Weleber, R. G. (2000). Mutations in NYX, encoding the leucine-rich proteoglycan nyctalopin, cause X-linked complete congenital stationary night blindness. *Nature Genetics*, 26(3), 319-23.

Bechynova, R., Dostál, J., Stratil, A., Jílek, F. & Horák, P. (2008). Mutation in the RPE65 gene causing hereditary retinal dystrophy in the Briard dogs: application of a new detection method. *Czech Journal of Animal Science*, 53(4), 176.

Bee, Y. M., Chawla, M. & Zhao, Y. (2015). Whole Exome Sequencing Identifies a Novel and a Recurrent Mutation in BBS2 Gene in a Family with Bardet-Biedl Syndrome. *BioMed Research International*, 2015, 524754.

Belancio, V. P., Deininger, P. L. & Roy-Engel, A. M. (2009). LINE dancing in the human genome: transposable elements and disease. *Genome Medicine*, 1(10), 97.

Beltran, W. A., Cideciyan, A. V., Lewin, A. S., Iwabe, S., Khanna, H., Sumaroka, A., Chiodo, V. A., Fajardo, D. S., Roman, A. J., Deng, W. T., Swider, M., Aleman, T. S., Boye, S. L., Genini, S., Swaroop, A., Hauswirth, W. W., Jacobson, S. G. & Aguirre, G. D. (2012). Gene therapy rescues photoreceptor blindness in dogs and paves the way for treating human X-linked retinitis pigmentosa. *Proceedings of the National Academy of Sciences of the United States of America,* 109(6), 2132-7.

Beltran, W. A., Cideciyan, A. V., Guziewicz, K. E., Iwabe, S., Swider, M., Scott, E. M., Savina, S. V., Ruthel, G., Stefano, F., Zhang, L., Zorger, R., Sumaroka, A., Jacobson, S. G. & Aguirre, G. D. (2014). Canine retina has a primate fovea-like bouquet of cone photoreceptors which is affected by inherited macular degenerations. *PLoS One*, 9(3), e90390.

Berkeley Drosophila Genome Project, 2020. https://www.fruitfly.org/seq_tools/splice.html. Accessed 16 March 2020.

Betts, M.J. and Russell, R.B. (2007). 'Amino-Acid Properties and Consequences of Substitutions' in *Bioinformatics for Geneticists*, M.R. Barnes (Ed.) pp. 289-316.

Bhattacharya, S. S., Wright, A. F., Clayton, J. F., Price, W. H., Phillips, C. I., McKeown, C. M., Jay, M., Bird, A. C., Pearson, P. L., Southern, E. M. & Evans, H. J. (1984). Close genetic linkage between X-

linked retinitis pigmentosa and a restriction fragment length polymorphism identified by recombinant DNA probe L1.28. *Nature*, 309(5965), 253-5.

Bibliowicz, J., Tittle, R. K. & Gross, J. M. (2011). Toward a better understanding of human eye disease insights from the zebrafish, Danio rerio. *Progress in molecular biology and translational science*, *100*, 287–330.

Bin, J., Madhavan, J., Ferrini, W., Mok, C. A., Billingsley, G., & Héon, E. (2009). BBS7 and TTC8 (BBS8) mutations play a minor role in the mutational load of Bardet-Biedl syndrome in a multiethnic population. *Human mutation*, *30*(7), E737–E746.

Blake, J. A., Eppig, J. T., Kadin, J. A., Richardson, J. E., Smith, C. L., Bult, C. J., & the Mouse Genome Database Group (2017). Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. *Nucleic acids research*, *45*(D1), D723–D729.

Borgström, M. K., Riise, R., Tornqvist, K. & Granath, L. (1996). Anomalies in the permanent dentition and other oral findings in 29 individuals with Laurence-Moon-Bardet-Biedl syndrome. *Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology*, 25(2), 86-89.

Botigué, L. R., Song, S., Scheu, A., Gopalan, S., Pendleton, A. L., Oetjens, M., Taravella, A. M., Seregély, T., Zeeb-Lanz, A., Arbogast, R. M., Bobo, D., Daly, K., Unterländer, M., Burger, J., Kidd, J. M., & Veeramah, K. R. (2017). Ancient European dog genomes reveal continuity since the Early Neolithic. *Nature communications*, 8, 16082.

Bowne, S. J., Humphries, M. M., Sullivan, L. S., Kenna, P. F., Tam, L. C., Kiang, A. S., Campbell, M., Weinstock, G. M., Koboldt, D. C., Ding, L., Fulton, R. S., Sodergren, E. J., Allman, D., Millington-Ward, S., Palfi, A., McKee, A., Blanton, S. H., Slifer, S., Konidari, I., Farrar, G. J., Daiger, S.P., Humphries, P. (2011). A dominant mutation in RPE65 identified by whole-exome sequencing causes retinitis pigmentosa with choroidal involvement. *European journal of human genetics: EJHG*, *19*(10), 1074–1081.

Boylan, J. P., & Wright, A. F. (2000). Identification of a novel protein interacting with RPGR. *Human molecular genetics*, 9(14), 2085–2093.

Brandl, C., Schulz, H. L., Charbel Issa, P., Birtel, J., Bergholz, R., Lange, C., Dahlke, C., Zobor, D., Weber, B. H. F. & Stohr, H. (2017). Mutations in the Genes for Interphotoreceptor Matrix Proteoglycans, IMPG1 and IMPG2, in Patients with Vitelliform Macular Lesions. *Genes (Basel)*, 8(7).

Breen, M., Bullerdiek, J. & Langford, C.F. (1999). The DAPI Banded Karyotype of the Domestic Dog (Canis familiaris) Generated Using Chromosome-Specific Paint Probes. *Chromosome Research*, 7, 401–406.

British Veterinary Association (2018). *2018 Eye examination results by breed*. Available online: https://www.bva.co.uk/media/2813/bva-kc-isds-eye-scheme-breed-specific-information-2018.pdf (Accessed: 16 March 2020).

Broeckx, B. J., Coopman, F., Verhoeven, G. E., Bavegems, V., De Keulenaer, S., De Meester, E., Van Niewerburgh, F. & Deforce, D. (2014). Development and performance of a targeted whole exome sequencing enrichment kit for the dog (Canis Familiaris Build 3.1). *Scientific reports*, 4, 5597.

Brooks, M. B., Gu, W., Barnas, J. L., Ray, J. & Ray, K. (2003). A Line 1 insertion in the Factor IX gene segregates with mild hemophilia B in dogs. *Mammalian genome: official journal of the International Mammalian Genome Society*, 14(11), 788-95.

Burns, M. E. & Arshavsky, V. Y. (2005). Beyond Counting Photons: Trials and Trends in Vertebrate Visual Transduction. *Neuron*, 48(3), 387-401.

Busse, C., Barnett, K. C., Mellersh, C. S. & Adams, V. J. (2011). Ophthalmic and cone derived electrodiagnostic findings in outbred Miniature Long-haired Dachshunds homozygous for a RPGRIP1 mutation. *Veterinary ophthalmology*, 14(3), 146-52.

Cancello, R., Tounian, A., Poitou, C. & Clément, K. (2004). Adiposity signals, genetic and body weight regulation in humans. *Diabetes & metabolism*, 30(3), 215-227.

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. & Werner, T. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21(13), 2933-42.

Castro-Sanchez, S., Alvarez-Satta, M., Tohamy, M. A., Beltran, S., Derdak, S. & Valverde, D. (2017). Whole exome sequencing as a diagnostic tool for patients with ciliopathy-like phenotypes. *PLoS One*, 12(8), e0183081.

Cehajic Kapetanovic, J., McClements, M. E., Martinez-Fernandez de la Camara, C., & MacLaren, R. E. (2019). Molecular Strategies for RPGR Gene Therapy. *Genes*, 10(9), 674.

Chaki, M., Airik, R., Ghosh, A. K., Giles, R. H., Chen, R., Slaats, G. G., Wang, H., Hurd, T. W., Zhou, W., Cluckey, A., Gee, H. Y., Ramaswami, G., Hong, C. J., Hamilton, B. A., Cervenka, I., Ganji, R. S., Bryja, V., Arts, H. H., van Reeuwijk, J., Oud, M. M., Letteboer, S. J., Roepman, R., Husson, H., Ibraghimov-Beskrovnaya, O., Yasunaga, T., Walz, G., Eley, L., Sayer, J. A., Schermer, B., Liebau, M. C., Benzing, T.,

Le Corre, S., Drummond, I., Janssen, S., Allen, S. J., Natarajan, S., O'Toole, J. F., Attanasio, M., Saunier, S., Antignac, C., Koenekoop, R. K., Ren, H., Lopez, I., Nayir, A., Stoetzel, C., Dollfus, H., Massoudi, R., Gleeson, J. G., Andreoli, S. P., Doherty, D. G., Lindstrad, A., Golzio, C., Katsanis, N., Pape, L., Abboud, E. B., Al-Rajhi, A. A., Lewis, R. A., Omran, H., Lee, E. Y., Wang, S., Sekiguchi, J. M., Saunders, R., Johnson, C. A., Garner, E., Vanselow, K., Andersen, J. S., Shlomai, J., Nurnberg, G., Nurnberg, P., Levy, S., Smogorzewska, A., Otto, E. A. & Hildebrandt, F. (2012). Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling. *Cell*, 150(3), 533-48.

Challis, B. G., Pritchard, L. E., Creemers, J. W. M., Delplanque, J., Keogh, J. M., Luan, J. a., Wareham, N. J., Yeo, G. S. H., Bhattacharyya, S., Froguel, P., White, A., Farooqi, I. S. & O'Rahilly, S. (2002). A missense mutation disrupting a dibasic prohormone processing site in pro-opiomelanocortin (POMC) increases susceptibility to early-onset obesity through a novel molecular mechanism. *Human molecular genetics*, 11(17), 1997-2004.

Chang, B., Khanna, H., Hawes, N., Jimeno, D., He, S., Lillo, C., Parapuram, S. K., Cheng, H., Scott, A., Hurd, R. E., Sayer, J. A., Otto, E. A., Attanasio, M., O'Toole, J. F., Jin, G., Shou, C., Hildebrandt, F., Williams, D. S., Heckenlively, J. R. & Swaroop, A. (2006). In-frame deletion in a novel centrosomal/ciliary protein CEP290/NPHP6 perturbs its interaction with RPGR and results in early-onset retinal degeneration in the rd16 mouse. *Human Molecular Genetics*, 15(11), 1847-57.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4, 7.

Chen, C. K., Burns, M. E., Spencer, M., Niemi, G. A., Chen, J., Hurley, J. B., Baylor, D. A. & Simon, M. I. (1999). Abnormal photoresponses and light-induced apoptosis in rods lacking rhodopsin kinase. *Proceedings of the National Academy of Sciences*, 96(7), 3718-3722.

Chen, J., Smaoui, N., Hammer, M. B. H., Jiao, X., Riazuddin, S. A., Harper, S., Katsanis, N., Riazuddin, S., Chaabouni, H., Berson, E. L. & Hejtmancik, J. F. (2011). Molecular Analysis of Bardet-Biedl Syndrome Families: Report of 21 Novel Mutations in 10 Genes. *Investigative Ophthalmology & Visual Science*, 52(8), 5317.

Chen, Q., Lee, J. W., Nishiyama, K., Shadrach, K. G., Rayborn, M. E. & Hollyfield, J. G. (2003). SPACRCAN in the interphotoreceptor matrix of the mouse retina: molecular, developmental and promoter analysis. *Experimental Eye Research*, 76(1), 1-14.

Cherry, T. J., Yang, M. G., Harmin, D. A., Tao, P., Timms, A. E., Bauwens, M., Allikmets, R., Jones, E. M., Chen, R., Baere, E. D. & Greenberg, M. E. (2018). Epigenomic Profiling and Single-Nucleus-RNA-

Seq Reveal *Cis*-Regulatory Elements in Human Retina, Macula and RPE and Non-Coding Genetic Variation. *bioRxiv*, 412361.

Chew, T., Haase, B., Bathgate, R., Willet, C. E., Kaukonen, M. K., Mascord, L. J., Lohi, H. T. & Wade, C. M. (2017). A Coding Variant in the Gene Bardet-Biedl Syndrome 4 (BBS4) Is Associated with a Novel Form of Canine Progressive Retinal Atrophy. *G3 (Bethesda)*, 7(7), 2327-2335.

Chiang, A. P., Nishimura, D., Searby, C., Elbedour, K., Carmi, R., Ferguson, A. L., Secrist, J., Braun, T., Casavant, T., Stone, E. M. & Sheffield, V. C. (2004). Comparative genomic analysis identifies an ADP-ribosylation factor-like gene as the cause of Bardet-Biedl syndrome (BBS3). *American Journal of Human Genetics*, 75(3), 475-484.

Chiang, A. P., Beck, J. S., Yen, H. J., Tayeh, M. K., Scheetz, T. E., Swiderski, R. E., Nishimura, D. Y., Braun, T. A., Kim, K. Y. A., Huang, J., Elbedour, K., Carmi, R., Slusarski, D. C., Casavant, T. L., Stone, E. M. & Sheffield, V. C. (2006). Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). *Proceedings of the National Academy of Sciences of the United States of America*, 103(16), 6287–6292.

Chinchore, Y., Mitra, A. & Dolph, P. J. (2009). Accumulation of rhodopsin in late endosomes triggers photoreceptor cell degeneration. *PLoS Genetics*, 5(2), e1000377.

Chuang, H.-C., Tsai, C.-Y., Hsueh, C.-H. & Tan, T.-H. (2018). GLK-IKKβ signaling induces dimerization and translocation of the AhR-RORγt complex in IL-17A induction and autoimmune disease. *Science Advances*, 4(9), eaat5401.

Clarke, G. M., Anderson, C. A., Pettersson, F. H., Cardon, L. R., Morris, A. P. & Zondervan, K. T. (2011). Basic statistical analysis in genetic case-control studies. *Nature Protocols*, 6(2), 121-33.

Clements, P. J., Gregory, C. Y., Peterson-Jones, S. M., Sargan, D. R. & Bhattacharya, S. S. (1993). Confirmation of the rod cGMP phosphodiesterase beta subunit (PDE beta) nonsense mutation in affected rcd-1 Irish setters in the UK and development of a diagnostic test. *Current Eye Research*, 12(9), 861-6.

CleverPet (2019). *Taking Care of Your Dog's Eyes.* Available online:

https://clever.pet/blogs/posts/taking-care-of-your-dogs-eye-health (Accessed: 01 March 2020).

Coene, K. L. M., Roepman, R., Doherty, D., Afroze, B., Kroes, H. Y., Letteboer, S. J. F., Ngu, L. H., Budny, B., Van Wijk, E., Gorden, N. T., Azhimi, M., Thauvin-Robinet, C., Veltman, J. A., Boink, M., Kleefstra, T., Cremers, F. P. M., Van Bokhoven, H. & De Brouwer, A. P. M. (2009). OFD1 is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin. *American Journal of Human Genetics*, 85(4), 465-481.

Courcier, E. A., Thomson, R. M., Mellor, D. J. & Yam, P. S. (2010). An epidemiological study of environmental factors associated with canine obesity. *Journal of Small Animal Practice*, 51(7), 362-367.

Credille, K. M., Minor, J. S., Barnhart, K. F., Lee, E., Cox, M. L., Tucker, K. A., Diegel, K. L., Venta, P. J., Hohl, D., Huber, M. & Dunstan, R. W. (2009). Transglutaminase 1-deficient recessive lamellar ichthyosis associated with a LINE-1 insertion in Jack Russell terrier dogs. *The British Journal of Dermatology*, 161(2), 265–272.

Curtis, R. & Barnett, K. C. (1993). Progressive retinal atrophy in miniature longhaired dachshund dogs. *The British Veterinary Journal*, 149(1), 71-85.

D'Cruz, P. M., Yasumura, D., Weir, J., Matthes, M. T., Abderrahim, H., LaVail, M. M. & Vollrath, D. (2000). Mutation of the receptor tyrosine kinase gene Mertk in the retinal dystrophic RCS rat. *Human Molecular Genetics*, 9(4), 645-51.

Davidson, A. E., Liskova, P., Evans, C. J., Dudakova, L., Noskova, L., Pontikos, N., Hartmannova, H., Hodanova, K., Stranecky, V., Kozmik, Z., Levis, H. J., Idigo, N., Sasai, N., Maher, G. J., Bellingham, J., Veli, N., Ebenezer, N. D., Cheetham, M. E., Daniels, J. T., Thaung, C. M., Jirsova, K., Plagnol, V., Filipec, M., Kmoch, S., Tuft, S. J. & Hardcastle, A. J. (2016). Autosomal-Dominant Corneal Endothelial Dystrophies CHED1 and PPCD1 Are Allelic Disorders Caused by Non-coding Mutations in the Promoter of OVOL2. *American Journal of Human Genetics*, 98(1), 75-89.

Davis, R. E., Swiderski, R. E., Rahmouni, K., Nishimura, D. Y., Mullins, R. F., Agassandian, K., Philp, A. R., Searby, C. C., Andrews, M. P., Thompson, S., Berry, C. J., Thedens, D. R., Yang, B., Weiss, R. M., Cassell, M. D., Stone, E. M. & Sheffield, V. C. (2007). A knockin mouse model of the Bardet Biedl syndrome 1 M390R mutation has cilia defects, ventriculomegaly, retinopathy, and obesity. *Proceedings of the National Academy of Sciences*, 104(49), 19422-19427.

de Vooght, K. M., van Wijk, R. & van Solinge, W. W. (2009). Management of gene promoter mutations in molecular diagnostics. *Clinical Chemistry*, 55(4), 698-708.

Deane, J. A., Verghese, E., Martelotto, L. G., Cain, J. E., Galtseva, A., Rosenblum, N. D., Watkins, D. N. & Ricardo, S. D. (2013). Visualizing renal primary cilia. *Nephrology (Carlton, Vic.)*, 18(3), 161–168.

Dekomien, G., Runte, M., Godde, R. & Epplen, J. T. (2000). Generalized progressive retinal atrophy of Sloughi dogs is due to an 8-bp insertion in exon 21 of the PDE6B gene. *Cytogenetics and Cell Genetics*, 90(3-4), 261-7.

Dekomien, G., Vollrath, C., Petrasch-Parwez, E., Boeve, M. H., Akkad, D. A., Gerding, W. M. & Epplen, J. T. (2010). Progressive retinal atrophy in Schapendoes dogs: mutation of the newly identified CCDC66 gene. *Neurogenetics*, 11(2), 163-74.

den Hollander, A. I., Koenekoop, R. K., Yzer, S., Lopez, I., Arends, M. L., Voesenek, K. E., Zonneveld, M. N., Strom, T. M., Meitinger, T., Brunner, H. G., Hoyng, C. B., van den Born, L. I., Rohrschneider, K. & Cremers, F. P. (2006). Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *American Journal of Human Genetics*, 79(3), 556-61.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D. & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491-8.

Dixon, C. J. (2016). Achromatopsia in three sibling Labrador Retrievers in the UK. *Veterinary Ophthalmology*, 19(1), 68-72.

Donders, F. C. (1857). Beiträge zur pathologischen Anatomie des Auges. *Albrecht von Graefes Archiv für Ophthalmologie*, 3(1), 139-165.

Dourlen, P., Bertin, B., Chatelain, G., Robin, M., Napoletano, F., Roux, M. J. & Mollereau, B. (2012). Drosophila fatty acid transport protein regulates rhodopsin-1 metabolism and is required for photoreceptor neuron survival. *PLoS Genetics*, 8(7), e1002833.

Dowling, J. E. & Sidman, R. L. (1962). Inherited retinal dystrophy in the rat. *The Journal of Cell Biology*, 14, 73-109.

Downs, L. M., Wallin-Hakansson, B., Boursnell, M., Marklund, S., Hedhammar, A., Truve, K., Hubinette, L., Lindblad-Toh, K., Bergstrom, T. & Mellersh, C. S. (2011). A frameshift mutation in golden retriever dogs with progressive retinal atrophy endorses SLC4A3 as a candidate gene for human retinal degenerations. *PLoS One*, 6(6), e21452.

Downs, L. M., Bell, J. S., Freeman, J., Hartley, C., Hayward, L. J. & Mellersh, C. S. (2013). Late-onset progressive retinal atrophy in the Gordon and Irish Setter breeds is associated with a frameshift mutation in C2orf71. *Animal Genetics*, 44(2), 169-77.

Downs, L. M. & Mellersh, C. S. (2014). An Intronic SINE insertion in FAM161A that causes exon-skipping is associated with progressive retinal atrophy in Tibetan Spaniels and Tibetan Terriers. *PLoS One*, 9(4), e93990.

Downs, L. M., Wallin-Hakansson, B., Bergstrom, T. & Mellersh, C. S. (2014). A novel mutation in TTC8 is associated with progressive retinal atrophy in the golden retriever. *Canine Genetics Epidemiology*, 1, 4.

Downs, L. M., Scott, E. M., Cideciyan, A. V., Iwabe, S., Dufour, V., Gardiner, K. L., Genini, S., Marinho, L. F., Sumaroka, A., Kosyk, M. S., Swider, M., Aguirre, G. K., Jacobson, S. G., Beltran, W. A. & Aguirre, G. D. (2016). Overlap of abnormal photoreceptor development and progressive degeneration in Leber congenital amaurosis caused by NPHP5 mutation. *Human Molecular Genetics*, 25(19), 4211-4226.

Dryja, T. P., McGee, T. L., Hahn, L. B., Cowley, G. S., Olsson, J. E., Reichel, E., Sandberg, M. A. & Berson, E. L. (1990). Mutations within the Rhodopsin Gene in Patients with Autosomal Dominant Retinitis Pigmentosa, *The New England Journal of Medicine,* 323(19), 1302-1307.

Ece Solmaz, A., Onay, H., Atik, T., Aykut, A., Cerrah Gunes, M., Ozalp Yuregir, O., Bas, V. N., Hazan, F., Kirbiyik, O. & Ozkinay, F. (2015). Targeted multi-gene panel testing for the diagnosis of Bardet Biedl syndrome: Identification of nine novel mutations across BBS1, BBS2, BBS4, BBS7, BBS9, BBS10 genes. *European Journal of Medical Genetics*, 58(12), 689-694.

Edney, A. T. & Smith, P. M. (1986). Study of obesity in dogs visiting veterinary practices in the United Kingdom. *The Veterinary Record*, 118(14), 391-396.

Ehret, G. B. (2010). Genome-Wide Association Studies: Contribution of Genomics to Understanding Blood Pressure and Essential Hypertension. *Current Hypertension Reports* 12(1), 17-25.

Eichers, E. R., Abd-El-Barr, M. M., Paylor, R., Lewis, R. A., Bi, W., Lin, X., Meehan, T. P., Stockton, D. W., Wu, S. M., Lindsay, E., Justice, M. J., Beales, P. L., Katsanis, N. & Lupski, J. R. (2006). Phenotypic characterization of Bbs4 null mice reveals age-dependent penetrance and variable expressivity. *Human Genetics,* 120(2), 211-226.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Erik, L., Hirsh, L., Paladin, L., Piovesan, D., Silvio, C. & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1), D427-D432.

EMBL-EBI (2020). *Ensembl Variation - Calculated variant consequences*. Available online: https://m.ensembl.org/info/genome/variation/prediction/predicted data.html (Accessed: 24 March 2020).

Everson, R., Pettitt, L., Forman, O. P., Dower-Tylee, O., McLaughlin, B., Ahonen, S., Kaukonen, M., Komaromy, A. M., Lohi, H., Mellersh, C. S., Sansom, J. & Ricketts, S. L. (2017). An intronic LINE-1

insertion in MERTK is strongly associated with retinopathy in Swedish Vallhund dogs. *PLoS One*, 12(8), e0183021.

Fadaie, Z., Khan, M., Del Pozo-Valero, M., Cornelis, S. S., Ayuso, C., Cremers, F., Roosing, S., & The Abca Study Group (2019). Identification of splice defects due to noncanonical splice site or deep-intronic variants in ABCA4. *Human mutation*, 40(12), 2365–2376.

Fain, G. L. (2006). Why photoreceptors die (and why they don't). *BioEssays : news and reviews in molecular, cellular and developmental biology*, 28(4), 344-354.

Fanning, T. G. & Singer, M. F. (1987). LINE-1: a mammalian transposable element. *Biochimica et biophysica acta*, 910(3), 203-212.

Farber, D. B. & Lolley, R. N. (1974). Cyclic guanosine monophosphate: elevation in degenerating photoreceptor cells of the C3H mouse retina. *Science*, 186(4162), 449-51.

Fédération Cynologique Internationale (2020). *Presentation of our organisation*. Available online: http://www.fci.be/en/Presentation-of-our-organisation-4.html (Accessed: 20 April 2020).

Ferreira do Amaral, C. O., Logar, G. d. A., Parisi, A. G., Takahashi, K. & Straioto, F. G. (2014). General and stomatologic aspects of bardet-biedl syndrome. *The Journal of Craniofacial Surgery*, 25(6), e575-e578.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews. Genetics*, 9(5), 397-405.

Feuillan, P. P., Ng, D., Han, J. C., Sapp, J. C., Wetsch, K., Spaulding, E., Zheng, Y. C., Caruso, R. C., Brooks, B. P., Johnston, J. J., Yanovski, J. A. & Biesecker, L. G. (2011). Patients with Bardet-Biedl Syndrome Have Hyperleptinemia Suggestive of Leptin Resistance. *The Journal of Clinical Endocrinology & Metabolism*, 96(3), E528-E535.

Fliegauf, M., Benzing, T. & Omran, H. (2007). When cilia go bad: cilia defects and ciliopathies. *Nature Reviews. Molecular Cell Biology*, 8(11), 880-93.

Forman, O. P., Pettitt, L., Komaromy, A. M., Bedford, P. & Mellersh, C. (2015). A Novel Genome-Wide Association Study Approach Using Genotyping by Exome Sequencing Leads to the Identification of a Primary Open Angle Glaucoma Associated Inversion Disrupting ADAMTS17. *PLoS One*, 10(12), e0143546.

Forman, O. P., Hitti, R. J., Boursnell, M., Miyadera, K., Sargan, D. & Mellersh, C. (2016). Canine genome assembly correction facilitates identification of a MAP9 deletion as a potential age of

onset modifier for RPGRIP1-associated canine retinal degeneration. *Mammalian genome: official journal of the International Mammalian Genome Society*, 27(5-6), 237-45.

Forsythe, E. & Beales, P. L. (2013). Bardet–Biedl syndrome. *European Journal of Human Genetics*, 21(1), 8-13.

Forsythe, E., Sparks, K., Best, S., Borrows, S., Hoskins, B., Sabir, A., Barrett, T., Williams, D., Mohammed, S., Goldsmith, D., Milford, D. V., Bockenhauer, D., Foggensteiner, L. & Beales, P. L. (2017). Risk Factors for Severe Renal Disease in Bardet–Biedl Syndrome. *Journal of the American Society of Nephrology*, 28(3), 963-970.

Forsythe, E., Kenny, J., Bacchelli, C. & Beales, P. L. (2018). Managing Bardet-Biedl Syndrome-Now and in the Future. *Frontiers in Pediatrics*, 6, 23.

Francisco, L. V., Langsten, A. A., Mellersh, C. S., Neal, C. L. & Ostrander, E. A. (1996). A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. *Mammalian genome : official journal of the International Mammalian Genome Society*, 7(5), 359-362.

Frank, V., den Hollander, A. I., Bruchle, N. O., Zonneveld, M. N., Nurnberg, G., Becker, C., Du Bois, G., Kendziorra, H., Roosing, S., Senderek, J., Nurnberg, P., Cremers, F. P., Zerres, K. & Bergmann, C. (2008). Mutations of the CEP290 gene encoding a centrosomal protein cause Meckel-Gruber syndrome. *Human Mutation*, 29(1), 45-52.

Fu, Q., Xu, M., Chen, X., Sheng, X., Yuan, Z., Liu, Y., Li, H., Sun, Z., Li, H., Yang, L., Wang, K., Zhang, F., Li, Y., Zhao, C., Sui, R. & Chen, R. (2017). CEP78 is mutated in a distinct type of Usher syndrome. *Journal of Medical Genetics*, 54(3), 190-195.

Gal, A., Schinzel, A., Orth, U., Fraser, N. A., Mollica, F., Craig, I. W., Kruse, T., Machler, M., Neugebauer, M. & Bleeker-Wagemakers, L. M. (1989). Gene of X-chromosomal congenital stationary night blindness is closely linked to DXS7 on Xp. *Human Genetics*, 81(4), 315-8.

Gal, A., Li, Y., Thompson, D. A., Weir, J., Orth, U., Jacobson, S. G., Apfelstedt-Sylla, E. & Vollrath, D. (2000). Mutations in MERTK, the human orthologue of the RCS rat retinal dystrophy gene, cause retinitis pigmentosa. *Nature Genetics*, 26(3), 270-1.

Gelatt, K. N., Gilger, B.C., Kern, T.J. (2013). *Veterinary ophthalmology / editor, Kirk N. Gelatt ; associate editors, Brian C. Gilger, Thomas J. Kern.*, *1*, 2 vols., 5th ed. edition. United States: Ames, Iowa : Wiley-Blackwell, 2013.

Genetic Alliance; District of Columbia Department of Health. *Understanding Genetics: A District of Columbia Guide for Patients and Health Professionals. Washington (DC): Genetic Alliance*; 2010 Feb

17. Appendix B, Classic Mendelian Genetics (Patterns of Inheritance) Available from: https://www.ncbi.nlm.nih.gov/books/NBK132145/.

Genomatix software suite (2017). *Gene2Promoter: Retrieval and analysis of promoters.* Available online: https://www.genomatix.de/online help/help eldorado/Gene2Promoter Intro.html (Accessed: 06 June 2017).

German, A. J., Woods, G. R. T., Holden, S. L., Brennan, L. & Burke, C. (2018). Dangerous trends in pet obesity. *Veterinary Record*, 182(1), 25.1-25.

Goldberg, A. F., Moritz, O. L. & Williams, D. S. (2016). Molecular basis for photoreceptor outer segment architecture. *Progress in Retinal and Eye Research*, 55, 52-81.

Goldstein, O., Zangerl, B., Pearce-Kelling, S., Sidjanin, D. J., Kijas, J. W., Felix, J., Acland, G. M. & Aguirre, G. D. (2006). Linkage disequilibrium mapping in domestic dog breeds narrows the progressive rod-cone degeneration interval and identifies ancestral disease-transmitting chromosome. *Genomics*, 88(5), 541-50.

Goldstein, O., Kukekova, A. V., Aguirre, G. D. & Acland, G. M. (2010). Exonic SINE insertion in STK38L causes canine early retinal degeneration (erd). *Genomics*, 96(6), 362-8.

Goldstein, O., Jordan, J. A., Aguirre, G. D. & Acland, G. M. (2013a). A non-stop S-antigen gene mutation is associated with late onset hereditary retinal degeneration in dogs. *Molecular Vision*, 19, 1871-84.

Goldstein, O., Mezey, J. G., Schweitzer, P. A., Boyko, A. R., Gao, C., Bustamante, C. D., Jordan, J. A., Aguirre, G. D. & Acland, G. M. (2013b). IQCB1 and PDE6B mutations cause similar early onset retinal degenerations in two closely related terrier dog breeds. *Investigative Ophthalmology & Visual Science*, 54(10), 7005-19.

Goyal, S., Jager, M., Robinson, P. N. & Vanita, V. (2016). Confirmation of TTC8 as a disease gene for nonsyndromic autosomal recessive retinitis pigmentosa (RP51). *Clinical Genetics*, 89(4), 454-460.

Gu, S. M., Thompson, D. A., Srikumari, C. R., Lorenz, B., Finckh, U., Nicoletti, A., Murthy, K. R., Rathmann, M., Kumaramanickavel, G., Denton, M. J. & Gal, A. (1997). Mutations in RPE65 cause autosomal recessive childhood-onset severe retinal dystrophy. *Nature Genetics*, 17(2), 194-7.

Guo, D.-F., Lin, Z., Wu, Y., Searby, C., Thedens, D. R., Richerson, G. B., Usachev, Y. M., Grobe, J. L., Sheffield, V. C. & Rahmouni, K. (2019). The BBSome in POMC and AgRP Neurons Is Necessary for Body Weight Regulation and Sorting of Metabolic Receptors. *Diabetes*, 68(8), 1591-1603.

Guziewicz, K. E., Zangerl, B., Lindauer, S. J., Mullins, R. F., Sandmeyer, L. S., Grahn, B. H., Stone, E. M., Acland, G. M. & Aguirre, G. D. (2007). Bestrophin gene mutations cause canine multifocal retinopathy: a novel animal model for best disease. *Investigative Ophthalmology & Visual Science*, 48(5), 1959-67.

Hagstrom, S. A., Duyao, M., North, M. A. & Li, T. (1999). Retinal degeneration in tulp1-/- mice: vesicular accumulation in the interphotoreceptor matrix. *Investigative Ophthalmology & Visual Science*, 40(12), 2795-802.

Hardcastle, A. J., David-Gray, Z. K., Jay, M., Bird, A. C. & Bhattacharya, S. S. (1997). Localization of CSNBX (CSNB4) between the retinitis pigmentosa loci RP2 and RP3 on proximal Xp. *Investigative Ophthalmology & Visual Science*, 38(13), 2750-5.

Hardcastle, A. J., Thiselton, D. L., Van Maldergem, L., Saha, B. K., Jay, M., Plant, C., Taylor, R., Bird, A. C. & Bhattacharya, S. (1999). Mutations in the RP2 Gene Cause Disease in 10% of Families with Familial X-Linked Retinitis Pigmentosa Assessed in This Study. *American Journal of Human Genetics*, 64(4), 1210-1215.

Hardcastle, A. J., Thiselton, D. L., Zito, I., Ebenezer, N., Mah, T. S., Gorin, M. B. & Bhattacharya, S. S. (2000). Evidence for a new locus for X-linked retinitis pigmentosa (RP23). *Investigative Ophthalmology & Visual Science*, 41(8), 2080-6.

Heon, E., Kim, G., Qin, S., Garrison, J. E., Tavares, E., Vincent, A., Nuangchamnong, N., Scott, C. A., Slusarski, D. C. & Sheffield, V. C. (2016). Mutations in C8ORF37 cause Bardet Biedl syndrome (BBS21). *Human Molecular Genetics*, 25(11), 2283-2294.

Hitti, R. J., Oliver, J. A. C., Schofield, E. C., Bauer, A., Kaukonen, M., Forman, O. P., Leeb, T., Lohi, H., Burmeister, L. M., Sargan, D. & Mellersh, C. S. (2019). Whole Genome Sequencing of Giant Schnauzer Dogs with Progressive Retinal Atrophy Establishes NECAP1 as a Novel Candidate Gene for Retinal Degeneration. *Genes (Basel)*, 10(5).

Hsu, Y., Garrison, J. E., Kim, G., Schmitz, A. R., Searby, C. C., Zhang, Q., Datta, P., Nishimura, D. Y., Seo, S. & Sheffield, V. C. (2017). BBSome function is required for both the morphogenesis and maintenance of the photoreceptor outer segment. *PLoS Genetics*, 13(10), e1007057.

Iakhine, R. (2004). Novel Dominant Rhodopsin Mutation Triggers Two Mechanisms of Retinal Degeneration and Photoreceptor Desensitization. *Journal of Neuroscience*, 24(10), 2516-2526.

Illumina (2017). *An Introduction to Next-Generation Sequencing Technology*. Available online: https://www.illumina.com/documents/products/illumina sequencing introduction.pdf (Accessed: 12 March 2020).

Imhoff, O., Marion, V., Stoetzel, C., Durand, M., Holder, M., Sigaudy, S., Sarda, P., Hamel, C. P., Brandt, C., Dollfus, H. & Moulin, B. (2011). Bardet-Biedl Syndrome: A Study of the Renal and Cardiovascular Phenotypes in a French Cohort. *Clinical Journal of the American Society of Nephrology*, 6(1), 22.

Inatani, M. & Tanihara, H. (2002). Proteoglycans in retina. *Progress in Retinal and Eye Research*, 21(5), 429-47.

Irion, U., Krauss, J. & Nusslein-Volhard, C. (2014). Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development*, 141(24), 4827-30.

Jacobs, G. H., Deegan, J. F., Crognale, M. A. & Fenwick, J. A. (1993). Photopigments of dogs and foxes and their implications for canid vision. *Visual Neuroscience*, 10(1), 173-180.

Jagannathan, V., Drögemüller, C., Leeb, T., & Dog Biomedical Variant Database Consortium (DBVDC) (2019). A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. *Animal Genetics*, 50(6), 695–704.

Jeanes, E. C., Oliver, J., Ricketts, S. L., Gould, D. J., & Mellersh, C. S. (2019). Glaucoma-causing ADAMTS17 mutations are also reproducibly associated with height in two domestic dog breeds: selection for short stature may have contributed to increased prevalence of glaucoma. *Canine genetics and epidemiology*, 6, 5.

Jin, H., May, M., Tranebjærg, L., Kendall, E., Fontán, G., Jackson, J., Subramony, S. H., Arena, F., Lubs, H., Smith, S., Stevenson, R., Schwartz, C. & Vetrie, D. (1996). A novel X–linked gene, DDP, shows mutations in families with deafness (DFN–1), dystonia, mental deficiency and blindness. *Nature Genetics*, 14(2), 177-180.

Jin, H. & Nachury, M. V. (2009). The BBSome. *Current biology : CB,* 19(12), R472-R473.

Jin, H., White, S. R., Shida, T., Schulz, S., Aguiar, M., Gygi, S. P., Bazan, J. F. & Nachury, M. V. (2010). The Conserved Bardet-Biedl Syndrome Proteins Assemble a Coat that Traffics Membrane Proteins to Cilia. *Cell*, 141(7), 1208-1219.

Johnson, R. C., Nelson, G. W., Troyer, J. L., Lautenberger, J. A., Kessing, B. D., Winkler, C. A. & O'Brien, S. J. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*, 11(1), 724.

Joo, J. W. J., Hormozdiari, F., Han, B. & Eskin, E. (2016). Multiple testing correction in linear mixed models. *Genome Biology*, 17(62).

Kametaka, S., Kametaka, A., Yonekura, S., Haruta, M., Takenoshita, S., Goto, S. & Waguri, S. (2012). AP-1 clathrin adaptor and CG8538/Aftiphilin are involved in Notch signaling during eye development in Drosophila melanogaster, *Journal of Cell Science*, 125(Pt 3), 634–648.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., Sabatti, C. & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4), 348-54.

Karlstetter, M., Sorusch, N., Caramoy, A., Dannhausen, K., Aslanidis, A., Fauser, S., Boesl, M. R., Nagel-Wolfrum, K., Tamm, E. R., Jagle, H., Stoehr, H., Wolfrum, U. & Langmann, T. (2014). Disruption of the retinitis pigmentosa 28 gene Fam161a in mice affects photoreceptor ciliary structure and leads to progressive retinal degeneration. *Human Molecular Genetics*, 23(19), 5197-210.

Katsanis, N., Beales, P. L., Woods, M. O., Lewis, R. A., Green, J. S., Parfrey, P. S., Ansley, S. J., Davidson, W. S. & Lupski, J. R. (2000). Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. *Nature Genetics*, 26(1), 67-70.

Katsanis, N. (2001). Triallelic Inheritance in Bardet-Biedl Syndrome, a Mendelian Recessive Disorder. *Science (New York, N.Y.)*, 293(5538), 2256-2259.

Kaukonen, M., Quintero, I. B., Mukarram, A. K., Hytönen, M. K., Holopainen, S., Wickström, K., Kyöstilä, K., Arumilli, M., Jalomäki, S., Daub, C. O., Kere, J., Lohi, H. & Do, G. A. C. (2020). A putative silencer variant in a spontaneous canine model of retinitis pigmentosa. *PLoS Genetics*, 16(3), e1008659-e1008659.

Kemp, C. M. & Jacobson, S. G. (1992). Rhodopsin levels in the central retinas of normal miniature poodles and those with progressive rod-cone degeneration. *Experimental Eye Research*, 54(6), 947-56.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, 12(6), 996-1006.

Khan, A. O., & Al Teneiji, A. M. (2019). Homozygous and heterozygous retinal phenotypes in families harbouring *IMPG2* mutations. *Ophthalmic genetics*, *40*(3), 247–251.

Khateb, S., Zelinger, L., Mizrahi-Meissonnier, L., Ayuso, C., Koenekoop, R. K., Laxer, U., Gross, M., Banin, E. & Sharon, D. (2014). A homozygous nonsense CEP250 mutation combined with a heterozygous nonsense C2orf71 mutation is associated with atypical Usher syndrome. *Journal of Medical Genetics*, 51(7), 460-9.

Kijas, J. W., Cideciyan, A. V., Aleman, T. S., Pianta, M. J., Pearce-Kelling, S. E., Miller, B. J., Jacobson, S. G., Aguirre, G. D. & Acland, G. M. (2002). Naturally occurring rhodopsin mutation in the dog causes retinal dysfunction and degeneration mimicking human dominant retinitis pigmentosa. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9), 6328-33.

Kijas, J. W., Miller, B. J., Pearce-Kelling, S. E., Aguirre, G. D. & Acland, G. M. (2003). Canine models of ocular disease: outcross breedings define a dominant disorder present in the English mastiff and bull mastiff dog breeds. *The Journal of Heredity,* 94(1), 27–30.

Kijas, J. W., Zangerl, B., Miller, B., Nelson, J., Kirkness, E. F., Aguirre, G. D. & Acland, G. M. (2004). Cloning of the canine ABCA4 gene and evaluation in canine cone-rod dystrophies and progressive retinal atrophies. *Molecular Vision*, 10, 223-32.

Kim, S. K., Shindo, A., Park, T. J., Oh, E. C., Ghosh, S., Gray, R. S., Lewis, R. A., Johnson, C. A., Attie-Bittach, T., Katsanis, N. & Wallingford, J. B. (2010). Planar Cell Polarity Acts Through Septins to Control Collective Cell Movement and Ciliogenesis. *Science (New York, N.Y.),* 329(5997), 1337-1340.

Koressaar, T. & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23(10), 1289-91.

Kropatsch, R., Petrasch-Parwez, E., Seelow, D., Schlichting, A., Gerding, W. M., Akkad, D. A., Epplen, J. T. & Dekomien, G. (2010). Generalized progressive retinal atrophy in the Irish Glen of Imaal Terrier is associated with a deletion in the ADAM9 gene. *Molecular and Cellular Probes*, 24(6), 357-363.

Kropatsch, R., Akkad, D. A., Frank, M., Rosenhagen, C., Altmüller, J., Nürnberg, P., Epplen, J. T. & Dekomien, G. (2016). A large deletion in RPGR causes XLPRA in Weimaraner dogs. *Canine Genetics and Epidemiology*, 3(1).

Kuehn, M. H. & Hageman, G. S. (1999). Molecular characterization and genomic mapping of human IPM 200, a second member of a novel family of proteoglycans. *Molecular Cell Biology Research Communications : MCBRC*, 2(2), 103-110.

Kuehn, M. H., Stone, E. M. & Hageman, G. S. (2001). Organization of the human IMPG2 gene and its evaluation as a candidate gene in age-related macular degeneration and other retinal degenerative disorders. *Investigative Ophthalmology & Visual Science*, 42(13), 3123-3129.

Kukekova, A. V., Goldstein, O., Johnson, J. L., Richardson, M. A., Pearce-Kelling, S. E., Swaroop, A., Friedman, J. S., Aguirre, G. D. & Acland, G. M. (2009). Canine RD3 mutation establishes rod-cone

dysplasia type 2 (rcd2) as ortholog of human and murine rd3. *Mammalian genome: official journal of the International Mammalian Genome Society*, 20(2), 109-123.

Kumar, P., Henikoff, S. & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7), 1073-1081.

Kumaran, N., Moore, A. T., Weleber, R. G. & Michaelides, M. (2017). Leber congenital amaurosis/early-onset severe retinal dystrophy: clinical features, molecular genetics and therapeutic interventions. *British Journal of Ophthalmology*, 101(9), 1147-1154.

Kwitek-Black, A. E., Carmi, R., Duyk, G. M., Buetow, K. H., Elbedour, K., Parvari, R., Yandava, C. N., Stone, E. M. & Sheffield, V. C. (1993). Linkage of Bardet-Biedl syndrome to chromosome 16q and evidence for non-allelic genetic heterogeneity. *Nature Genetics*, 5(4), 392-396.

Kühnen, P., Clément, K., Wiegand, S., Blankenstein, O., Gottesdiener, K., Martini, L. L., Mai, K., Blume-Peytavi, U., Grüters, A. & Krude, H. (2016) Proopiomelanocortin Deficiency Treated with a Melanocortin-4 Receptor Agonist. *New England Journal of Medicine*, 375(3), 240-246.

La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., & Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, 352(6330), 77–79.

Lai, C. L., L'Eplattenier, H., van den Ham, R., Verseijden, F., Jagtenberg, A., Mol, J. A., & Teske, E. (2008). Androgen receptor CAG repeat polymorphisms in canine prostate cancer. *Journal of veterinary internal medicine*, 22(6), 1380–1384.

Lazarus, H. S. & Hageman, G. S. (1992) Xyloside-induced disruption of interphotoreceptor matrix proteoglycans results in retinal detachment. *Investigative Ophthalmology & Visual Science*, 33(2), 364-376.

Lee, Y. S., Challis, B. G., Thompson, D. A., Yeo, G. S. H., Keogh, J. M., Madonna, M. E., Wraight, V., Sims, M., Vatin, V., Meyre, D., Shield, J., Burren, C., Ibrahim, Z., Cheetham, T., Swift, P., Blackwood, A., Hung, C.-C. C., Wareham, N. J., Froguel, P., Millhauser, G. L., O'Rahilly, S. & Farooqi, I. S. (2006) A POMC variant implicates beta-melanocyte-stimulating hormone in the control of human energy balance. *Cell Metabolism*, 3(2), 135-140.

Lee, K. A., Nawrot, M., Garwin, G. G., Saari, J. C., & Hurley, J. B. (2010). Relationships among visual cycle retinoids, rhodopsin phosphorylation, and phototransduction in mouse eyes during light and dark adaptation. *Biochemistry*, 49(11), 2454–2463.

Leitch, C. C., Zaghloul, N. A., Davis, E. E., Stoetzel, C., Diaz-Font, A., Rix, S., Alfadhel, M., Lewis, R. A., Eyaid, W., Banin, E., Dollfus, H., Beales, P. L., Badano, J. L. & Katsanis, N. (2008). Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. *Nature Genetics*, 40(4), 443-448.

Lem, J., Krasnoperova, N. V., Calvert, P. D., Kosaras, B., Cameron, D. A., Nicolo, M., Makino, C. L. & Sidman, R. L. (1999). Morphological, physiological, and biochemical changes in rhodopsin knockout mice. *Proceedings of the National Academy of Sciences of the United States of America*, 96(2), 736-741.

Lheriteau, E., Petit, L., Weber, M., Le Meur, G., Deschamps, J. Y., Libeau, L., Mendes-Madeira, A., Guihal, C., Francois, A., Guyon, R., Provost, N., Lemoine, F., Papal, S., El-Amraoui, A., Colle, M. A., Moullier, P. & Rolling, F. (2014). Successful gene therapy in the RPGRIP1-deficient dog: a large model of cone-rod dystrophy. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, 22(2), 265-277.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and 1000 Genome Project Data Processing Subgroup (2009). The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25(16) 2078-9.

Li, H. & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), 589-595.

Li, J. B., Gerdes, J. M., Haycraft, C. J., Fan, Y., Teslovich, T. M., May-Simera, H., Li, H., Blacque, O. E., Li, L., Leitch, C. C., Lewis, R. A., Green, J. S., Parfrey, P. S., Leroux, M. R., Davidson, W. S., Beales, P. L., Guay-Woodford, L. M., Yoder, B. K., Stormo, G. D., Katsanis, N. & Dutcher, S. K. (2004). Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell*, 117(4), 541-552.

Li, T., Snyder, W. K., Olsson, J. E. & Dryja, T. P. (1996). Transgenic mice carrying the dominant rhodopsin mutation P347S: evidence for defective vectorial transport of rhodopsin to the outer segments. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24), 14176-14181.

Li, X., Chen, S., Wang, Q., Zack, D. J., Snyder, S. H. & Borjigin, J. (1998). A pineal regulatory element (PIRE) mediates transactivation by the pineal/retina-specific transcription factor CRX. *Proceedings of the National Academy of Sciences of the United States of America*, 95(4), 1876-1881.

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., 3rd, Zody, M. C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., DeJong, P. J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C. W., Cook, A., Cuff, J., Daly, M. J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K. P., Parker, H. G., Pollinger, J. P., Searle, S. M., Sutter, N. B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., et al (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438(7069), 803-819.

Lindstrand, A., Frangakis, S., Claudia, Ellen, Kelsey, Jason, Pehlivan, D., Liu, P., Igor, Sabo, A., Richard, Banin, E., James, Erica & Katsanis, N. (2016). Copy-Number Variation Contributes to the Mutational Load of Bardet-Biedl Syndrome. *American Journal of Human Genetics,* 99(2), 318-336.

Lingaas, F., Sorensen, A., Juneja, R. K., Johansson, S., Fredholm, M., Wintero, A. K., Sampson, J., Mellersh, C., Curzon, A., Holmes, N. G., Binns, M. M., Dickens, H. F., Ryder, E. J., Gerlach, J., Bäumle, E., & Dolf, G. (1997). Towards construction of a canine linkage map: establishment of 16 linkage groups. *Mammalian genome:official journal of the International Mammalian Genome Society*, 8(3), 218–221.

Livak, K. J. (1999). Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genetic Analysis: Biomolecular Engineering,* 14(5-6), 143-149.

Loktev, A. V., Zhang, Q., Beck, J. S., Searby, C. C., Scheetz, T. E., Bazan, J. F., Slusarski, D. C., Sheffield, V. C., Jackson, P. K. & Nachury, M. V. (2008). A BBSome Subunit Links Ciliogenesis, Microtubule Stability, and Acetylation. *Developmental Cell*, 15(6), 854-865.

Lorda-Sanchez, I., Ayuso, C., Sanz, R. & Ibañez, A. (2001). Does Bardet-Biedl syndrome have a characteristic face? *Journal of Medical Genetics*, 38(5), E14-E14.

Ludlam, W. G., Aoba, T., Cuellar, J., Bueno-Carrasco, M. T., Makaju, A., Moody, J. D., Franklin, S., Valpuesta, J. M. & Willardson, B. M. (2019). Molecular architecture of the Bardet-Biedl syndrome protein 2-7-9 subcomplex. *The Journal of Biological Chemistry,* 294(44), 16385–16399.

Lund, E. M., Armstrong, P. J., Kirk, C. A. & Klausner, J. S. (2006). Prevalence and Risk Factors for Obesity in Adult Dogs from Private US Veterinary Practices. *The International Journal Of Applied Research*, 4(2), 177-186.

Magnusson, H. (1911). Über Retinitis Pigmentosa und Konsanquinität Beim Hunde. *Archiv Für Vergleichende Ophthalmologie*, 2, 147-163.

Makelainen, S., Godia, M., Hellsand, M., Viluma, A., Hahn, D., Makdoumi, K., Zeiss, C. J., Mellersh, C., Ricketts, S. L., Narfström, K., Hallbook, F., Ekesten, B., Andersson, G. & Bergstrom, T. F. (2019). An ABCA4 loss-of-function mutation causes a canine form of Stargardt disease. *PLoS Genetics*, 15(3), e1007873.

Mannu, G. S. (2014). Retinal phototransduction. *Neurosciences (Riyadh)*, 19(4), 275-80.

Marion, V., Stutzmann, F., Gérard, M., De Melo, C., Schaefer, E., Claussmann, A., Hellé, S., Delague, V., Souied, E., Barrey, C., Verloes, A., Stoetzel, C. & Dollfus, H. (2012). Exome sequencing identifies mutations in LZTFL1, a BBSome and smoothened trafficking regulator, in a family with Bardet--Biedl syndrome with situs inversus and insertional polydactyly. *Journal of Medical Genetics*, 49(5), 317-321.

Marlhens, F., Bareil, C., Griffoin, J. M., Zrenner, E., Amalric, P., Eliaou, C., Liu, S. Y., Harris, E., Redmond, T. M., Arnaud, B., Claustres, M. & Hamel, C. P. (1997). Mutations in RPE65 cause Leber's congenital amaurosis. *Nature Genetics*, 17(2), 139-41.

Mattera, R., Ritter, B., Sidhu, S. S., McPherson, P. S., & Bonifacino, J. S. (2004). Definition of the consensus motif recognized by gamma-adaptin ear domains. *The Journal of Biological Chemistry*, 279(9), 8018–8028.

McGreevy, P., Grassi, T. D. & Harman, A. M. (2004). A Strong Correlation Exists between the Distribution of Retinal Ganglion Cells and Nose Length in the Dog. *Brain, Behavior and Evolution*, 63(1), 13–22.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297-1303.

McMahon, H. T. & Boucrot, E. (2011). Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nature Reviews Molecular Cell Biology*, 12(8), 517-533.

McWilliam, P., Farrar, G. J., Kenna, P., Bradley, D. G., Humphries, M. M., Sharp, E. M., McConnell, D. J., Lawler, M., Sheils, D., Ryan, C., Stevens, K., Daiger, S. P. & Humphries, P. (1989). Autosomal dominant retinitis pigmentosa (ADRP): Localization of an ADRP gene to the long arm of chromosome 3. *Genomics*, 5(3), 619-622.

Mellersh, C. S., Langston, A. A., Acland, G. M., Fleming, M. A., Ray, K., Wiegand, N. A., Francisco, L. V., Gibbs, M., Aguirre, G. D., & Ostrander, E. A. (1997). A linkage map of the canine genome. *Genomics*, 46(3), 326–336.

Mellersh, C. S., Boursnell, M. E., Pettitt, L., Ryder, E. J., Holmes, N. G., Grafham, D., Forman, O. P., Sampson, J., Barnett, K. C., Blanton, S., Binns, M. M. & Vaudin, M. (2006). Canine RPGRIP1 mutation establishes cone-rod dystrophy in miniature longhaired dachshunds as a homologue of human Leber congenital amaurosis. *Genomics*, 88(3), 293-301.

Mellersh, C. (2008). Give a dog a genome. *Veterinary Journal*, 178(1), 46-52.

Mellersh, C. S. (2014). The genetics of eye disorders in the dog. *Canine Genetics Epidemiology*, 1, 3.

Menotti-Raymond, M., David, V. A., Schaffer, A. A., Stephens, R., Wells, D., Kumar-Singh, R., O'Brien, S. J. & Narfström, K. (2007). Mutation in CEP290 discovered for cat model of human retinal degeneration. *The Journal of Heredity*, 98(3), 211-220.

Millichamp, N. J., Curtis, R. & Barnett, K. C. (1988). Progressive retinal atrophy in Tibetan terriers. *Journal of the American Veterinary Medical Association*, 192(6), 769-776.

Miyadera, K., Kato, K., Aguirre-Hernandez, J., Tokuriki, T., Morimoto, K., Busse, C., Barnett, K., Holmes, N., Ogawa, H., Sasaki, N., Mellersh, C. S. & Sargan, D. R. (2009). Phenotypic variation and genotype-phenotype discordance in canine cone-rod dystrophy with an RPGRIP1 mutation. *Molecular Vision*, 15, 2287-2305.

Miyadera, K., Kato, K., Boursnell, M., Mellersh, C. S. & Sargan, D. R. (2012). Genome-wide association study in RPGRIP1(-/-) dogs identifies a modifier locus that determines the onset of retinal degeneration. *Mammalian Genome: official journal of the International Mammalian Genome Society*, 23(1-2), 212-223.

Moaven, H., Koike, Y., Jao, C. C., Gurevich, V. V., Langen, R. & Chen, J. (2013). Visual arrestin interaction with clathrin adaptor AP-2 regulates photoreceptor survival in the vertebrate retina. *Proceedings of the National Academy of Sciences of the United States of America,* 110(23), 9463-9468.

Moiseyev, G., Chen, Y., Takahashi, Y., Wu, B. X. & Ma, J. X. (2005). RPE65 is the isomerohydrolase in the retinoid visual cycle. *Proceedings of the National Academy of Sciences of the United States of America,* 102(35), 12413-12418.

Moore, S. J., Green, J. S., Fan, Y., Bhogal, A. K., Dicks, E., Fernandez, B. A., Stefanelli, M., Murphy, C., Cramer, B. C., Dean, J. C. S., Beales, P. L., Katsanis, N., Bassett, A. S., Davidson, W. S. & Parfrey, P. S. (2005). Clinical and genetic epidemiology of Bardet-Biedl syndrome in Newfoundland: A 22-year prospective, population-based, cohort study. *American Journal of Medical Genetics*. Part A, 132A(4), 352–360.

Morton, G. J., Cummings, D. E., Baskin, D. G., Barsh, G. S. & Schwartz, M. W. (2006). Central nervous system control of food intake and body weight. *Nature*, 443(7109), 289-295.

Mowat, F. M., Occelli, L. M., Bartoe, J. T., Gervais, K. J., Bruewer, A. R., Querubin, J., Dinculescu, A., Boye, S. L., Hauswirth, W. W. & Petersen-Jones, S. M. (2017). Gene Therapy in a Large Animal Model of PDE6A-Retinitis Pigmentosa. *Frontiers in Neuroscience*, 11, 342.

Murgiano, L., Becker, D., Torjman, D., Niggel, J. K., Milano, A., Cullen, C., Feng, R., Wang, F., Jagannathan, V., Pearce-Kelling, S., Katz, M. L., Leeb, T. & Aguirre, G. D. (2018). Complex Structural PPT1 Variant Associated with Non-syndromic Canine Retinal Degeneration. *G3 (Bethesda)*.

Mykytyn, K., Braun, T., Carmi, R., Haider, N. B., Searby, C. C., Shastri, M., Beck, G., Wright, A. F., Iannaccone, A., Elbedour, K., Riise, R., Baldi, A., Raas-Rothschild, A., Gorman, S. W., Duhl, D. M., Jacobson, S. G., Casavant, T., Stone, E. M. & Sheffield, V. C. (2001). Identification of the gene that, when mutated, causes the human obesity syndrome BBS4. *Nature Genetics*, 28(2), 188-191.

Mykytyn, K., Nishimura, D. Y., Searby, C. C., Shastri, M., Yen, H.-j., Beck, J. S., Braun, T., Streb, L. M., Cornier, A. S., Cox, G. F., Fulton, A. B., Carmi, R., Lüleci, G., Chandrasekharappa, S. C., Collins, F. S., Jacobson, S. G., Heckenlively, J. R., Weleber, R. G., Stone, E. M. & Sheffield, V. C. (2002). Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nature Genetics*, 31(4), 435-438.

Mykytyn, K., Mullins, R. F., Andrews, M., Chiang, A. P., Swiderski, R. E., Yang, B., Braun, T., Casavant, T., Stone, E. M. & Sheffield, V. C. (2004). Bardet-Biedl syndrome type 4 (BBS4)-null mice implicate Bbs4 in flagella formation but not global cilia assembly. *Proceedings of the National Academy of Sciences of the United States of America,* 101(23), 8664-8669.

Nachury, M. V., Loktev, A. V., Zhang, Q., Westlake, C. J., Peranen, J., Merdes, A., Slusarski, D. C., Scheller, R. H., Bazan, J. F., Sheffield, V. C. & Jackson, P. K. (2007). A core complex of BBS proteins

cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell*, 129(6), 1201-1213.

Namburi, P., Ratnapriya, R., Khateb, S., Lazar, C. H., Kinarty, Y., Obolensky, A., Erdinest, I., Marks-Ohana, D., Pras, E., Ben-Yosef, T., Newman, H., Gross, M., Swaroop, A., Banin, E. & Sharon, D. (2016). Bi-allelic Truncating Mutations in CEP78, Encoding Centrosomal Protein 78, Cause Cone-Rod Degeneration with Sensorineural Hearing Loss. *American Journal of Human Genetics*, 99(3), 777-784.

Narfström, K., Wrigstad, A. & Nilsson, S. E. (1989). The Briard dog: a new animal model of congenital stationary night blindness. *Br J Ophthalmol*, 73(9), 750-756.

Narfström, K. (1999). Retinal dystrophy or 'congenital stationary night blindness' in the Briard dog. *Veterinary Ophthalmology*, 2(1), 75-76.

NCBI Mutation Analyzer (2017). Available online:

https://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa explorer.cgi?mode=translate (Accessed: 06 January 2020).

Neff, M. W., Broman, K. W., Mellersh, C. S., Ray, K., Acland, G. M., Aguirre, G. D., Ziegle, J. S., Ostrander, E. A., & Rine, J. (1999). A second-generation genetic linkage map of the domestic dog, Canis familiaris. *Genetics*, 151(2), 803–820.

Neitz, J., Geist, T. & Jacobs, G. H. (1989). Color vision in the dog. *Visual Neuroscience*, 3(2), 119-125.

Nikopoulos, K., Farinelli, P., Giangreco, B., Tsika, C., Royer-Bertrand, B., Mbefo, M. K., Bedoni, N., Kjellstrom, U., El Zaoui, I., Di Gioia, S. A., Balzano, S., Cisarova, K., Messina, A., Decembrini, S., Plainis, S., Blazaki, S. V., Khan, M. I., Micheal, S., Boldt, K., Ueffing, M., Moulin, A. P., Cremers, F. P. M., Roepman, R., Arsenijevic, Y., Tsilimbaris, M. K., Andreasson, S. & Rivolta, C. (2016). Mutations in CEP78 Cause Cone-Rod Dystrophy and Hearing Loss Associated with Primary-Cilia Defects. *American Journal of Human Genetics*, 99(3), 770-776.

Nishimura, D. Y., Searby, C. C., Carmi, R., Elbedour, K., Van Maldergem, L., Fulton, A. B., Lam, B. L., Powell, B. R., Swiderski, R. E., Bugge, K. E., Haider, N. B., Kwitek-Black, A. E., Ying, L., Duhl, D. M., Gorman, S. W., Heon, E., Iannaccone, A., Bonneau, D., Biesecker, L. G., Jacobson, S. G., Stone, E. M. & Sheffield, V. C. (2001). Positional cloning of a novel gene on chromosome 16q causing Bardet-Biedl syndrome (BBS2). *Human Molecular Genetics*, 10(8), 865-874.

Nishimura, D. Y., Fath, M., Mullins, R. F., Searby, C., Andrews, M., Davis, R., Andorf, J. L., Mykytyn, K., Swiderski, R. E., Yang, B., Carmi, R., Stone, E. M. & Sheffield, V. C. (2004). Bbs2-null mice have neurosensory deficits, a defect in social dominance, and retinopathy associated with

mislocalization of rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America*, 101(47), 16588-16593.

Nishimura, D. Y., Swiderski, R. E., Searby, C. C., Berg, E. M., Ferguson, A. L., Hennekam, R., Merin, S., Weleber, R. G., Biesecker, L. G., Stone, E. M. & Sheffield, V. C. (2005). Comparative Genomics and Gene Expression Analysis Identifies BBS9, a New Bardet-Biedl Syndrome Gene. *American Journal of Human Genetics*, 77(6), 1021-1033.

Occelli, L. M., Schon, C., Seeliger, M. W., Biel, M., Michalakis, S., Petersen-Jones, S. & Consortium, R. C. (2017). Gene Supplementation Rescues Rod Function and Preserves Photoreceptor and Retinal Morphology in Dogs, Leading the Way Towards Treating Human PDE6A-Retinitis Pigmentosa. *Hum Gene Therapy*.

Oliver, J., Rustidge, S., Pettitt, L., Jenkins, C. A., Farias, F., Giuliano, E. A., & Mellersh, C. S. (2018). Evaluation of ADAMTS17 in Chinese Shar-Pei with primary open-angle glaucoma, primary lens luxation, or both. *American journal of veterinary research*, 79(1), 98–106.

Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). (2020). World Wide Web URL: https://omim.org/

Ostertag, E. M. & Kazazian, H. H., Jr. (2001). Biology of mammalian L1 retrotransposons. *Annual Review of Genetics*, 35, 501-38.

Otto, E. A., Hurd, T. W., Airik, R., Chaki, M., Zhou, W., Stoetzel, C., Patil, S. B., Levy, S., Ghosh, A. K., Murga-Zamalloa, C. A., Van Reeuwijk, J., Letteboer, S. J. F., Sang, L., Giles, R. H., Liu, Q., Coene, K. L. M., Estrada-Cuzcano, A., Collin, R. W. J., McLaughlin, H. M., Held, S., Kasanuki, J. M., Ramaswami, G., Conte, J., Lopez, I., Washburn, J., Macdonald, J., Hu, J., Yamashita, Y., Maher, E. R., Guay-Woodford, L. M., Neumann, H. P. H., Obermüller, N., Koenekoop, R. K., Bergmann, C., Bei, X., Lewis, R. A., Katsanis, N., Lopes, V., Williams, D. S., Lyons, R. H., Dang, C. V., Brito, D. A., Dias, M. B., Zhang, X., Cavalcoli, J. D., Nürnberg, G., Nürnberg, P., Pierce, E. A., Jackson, P. K., Antignac, C., Saunier, S., Roepman, R., Dollfus, H., Khanna, H. & Hildebrandt, F. (2010). Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nature Genetics,* 42(10), 840-850.

Paez, G. L., Sellers, K. F., Band, M., Acland, G. M., Zangerl, B. & Aguirre, G. D. (2006). Characterization of gene expression profiles of normal canine retina and brain using a retinal cDNA microarray. *Molecular Vision*, 12, 1048-1056.

Palmer, E. E., Jarrett, K. E., Sachdev, R. K., Al Zahrani, F., Hashem, M. O., Ibrahim, N., Sampaio, H., Kandula, T., Macintosh, R., Gupta, R., Conlon, D. M., Billheimer, J. T., Rader, D. J., Funato, K., Walkey, C. J., Lee, C. S., Loo, C., Brammah, S., Elakis, G., Zhu, Y., Buckley, M., Kirk, E. P., Bye, A., Alkuraya, F.

S., Roscioli, T. & Lagor, W. R. (2016). Neuronal deficiency of ARV1 causes an autosomal recessive epileptic encephalopathy. *Human Molecular Genetics*, 25(14), 3042-3054.

Panny, A., Glurich, I., Haws, R. M. & Acharya, A. (2017). Oral and Craniofacial Anomalies of Bardet-Biedl Syndrome: Dental Management in the Context of a Rare Disease. J*ournal of Dental Research*, 96(12), 1361-1369.

Paquet-Durand, F., Beck, S., Michalakis, S., Goldmann, T., Huber, G., Mühlfriedel, R., Trifunović, D., Fischer, M. D., Fahl, E., Duetsch, G., Becirovic, E., Wolfrum, U., Van Veen, T., Biel, M., Tanimoto, N. & Seeliger, M. W. (2011). A key role for cyclic nucleotide gated (CNG) channels in cGMP-related retinitis pigmentosa. *Human Molecular Genetics*, 20(5), 941-947.

Park, S. J., Lee, D. S., Lim, E. J., Choi, S. H., Kang, W. S., Kim, I. B. & Chun, M. H. (2008). The absence of the clathrin-dependent endocytosis in rod bipolar cells of the FVB/N mouse retina. *Neuroscience Letters*, 439(2), 165-169.

Parker, H. G., Kim, L. V., Sutter, N. B., Carlson, S., Lorentzen, T. D., Malek, T. B., Johnson, G. S., DeFrance, H. B., Ostrander, E. A. & Kruglyak, L. (2004). Genetic structure of the purebred domestic dog. *Science*, 304(5674), 1160-1164.

Parker, H. G., Kukekova, A. V., Akey, D. T., Goldstein, O., Kirkness, E. F., Baysac, K. C., Mosher, D. S., Aguirre, G. D., Acland, G. M. & Ostrander, E. A. (2007). Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Research*, 17(11), 1562-1571.

Parker, H. G., Dreger, D. L., Rimbault, M., Davis, B. W., Mullen, A. B., Carpintero-Ramirez, G. & Ostrander, E. A. (2017). Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Reports*, 19(4), 697-708.

Parry H. B. (1953). Degenerations of the dog retina. II. Generalized progressive atrophy of hereditary origin. *The British Journal of Ophthalmology*, 37(8), 487–502.

Parshall, C. J., Wyman, M., Nitroy, S., Acland, G. M. & Aguirre, G. D. (1991). Photoreceptor dysplasia: an inherited progressive retinal atrophy of miniature schnauzer dogs. *Progress in Veterinary & Comparative Ophthalmology*, 1(3), 187.

Pedersen, N. C., Brucker, L., Tessier, N. G., Liu, H., Penedo, M. C., Hughes, S., Oberbauer, A. & Sacks, B. (2015). The effect of genetic bottlenecks and inbreeding on the incidence of two major autoimmune diseases in standard poodles, sebaceous adenitis and Addison's disease. *Canine Genetics Epidemiology*, 2, 14.

Peichl, L. (1992a). Morphological types of ganglion cells in the dog and wolf retina. *The Journal of Comparative Neurology*, 324(4), 590-602.

Peichl, L. (1992b). Topography of ganglion cells in the dog and wolf retina. *The Journal of Comparative Neurology*, 324(4), 603-620.

Penzkofer, T., Jäger, M., Figlerowicz, M., Badge, R., Mundlos, S., Robinson, P. N. & Zemojtel, T. (2017). L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Research*, 45(D1), D68–D73.

Petersen-Jones, S. M., Entz, D. D. & Sargan, D. R. (1999). cGMP phosphodiesterase-alpha mutation causes progressive retinal atrophy in the Cardigan Welsh corgi dog. *Investigative Ophthalmology & Visual Science*, 40(8), 1637-44.

Petersen-Jones S. (2005). Advances in the molecular understanding of canine retinal diseases. *The Journal of Small Animal Practice*, 46(8), 371–380.

Petit, L., Lheriteau, E., Weber, M., Le Meur, G., Deschamps, J. Y., Provost, N., Mendes-Madeira, A., Libeau, L., Guihal, C., Colle, M. A., Moullier, P. & Rolling, F. (2012). Restoration of vision in the pde6beta-deficient dog, a large animal model of rod-cone dystrophy. *Molecular Therapy : the Journal of the American Society of Gene Therapy*, 20(11), 2019-2030.

PFMA (2019). *Paw PDSA Animal Wellbeing Report 2019*. Available online: https://www.pdsa.org.uk/media/7420/2019-paw-report downloadable.pdf (Accessed: 06 January 2020).

Pierce, E. A. (2001). Pathways to photoreceptor cell death in inherited retinal degenerations. *Bioessays*, 23(7), 605-618.

Pittler, S. J. & Baehr, W. (1991). Identification of a nonsense mutation in the rod photoreceptor cGMP phosphodiesterase beta-subunit gene of the rd mouse. *Proceedings of the National Academy of Sciences of the United States of America*, 88(19), 8322-8326.

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G. & Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559-575.

Purves, D., Augustine, G.J., Fitzpatrick, D., Katz, L.C., LaMantia,A., McNamara, J.O., and Williams, S.M. (Eds) (2001). *Neuroscience. 2nd edition.*, 2nd edition. Sunderland (MA): Sinauer Associates.

Quignon, P., Herbin, L., Cadieu, E., Kirkness, E. F., Hédan, B., Mosher, D. S., Galibert, F., André, C., Ostrander, E. A., & Hitte, C. (2007). Canine population structure: assessment and impact of intra-breed stratification on SNP-based association studies. *PloS one*, 2(12), e1324.

Quinlan, A. R. & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842.

Raffan, E., Dennis, R. J., O'Donovan, C. J., Becker, J. M., Scott, R. A., Smith, S. P., Withers, D. J., Wood, C. J., Conci, E., Clements, D. N., Summers, K. M., German, A. J., Mellersh, C. S., Arendt, M. L., Iyemere, V. P., Withers, E., Söder, J., Wernersson, S., Andersson, G., Lindblad-Toh, K., Yeo, G.S., O'Rahilly, S. (2016). A Deletion in the Canine POMC Gene Is Associated with Weight and Appetite in Obesity-Prone Labrador Retriever Dogs. *Cell Metabolism*, 23(5), 893–900.

Rahmouni, K., Fath, M. A., Seo, S., Thedens, D. R., Berry, C. J., Weiss, R., Nishimura, D. Y. & Sheffield, V. C. (2008). Leptin resistance contributes to obesity and hypertension in mouse models of Bardet-Biedl syndrome. *The Journal of Clinical Investigation*, 118(4), 1458-1467.

Reddy, B., Patel, A. K., Singh, K. M., Patil, D. B., Parikh, P. V., Kelawala, D. N., Koringa, P. G., Bhatt, V. D., Rao, M. V. & Joshi, C. G. (2015). De Novo Assembly and Transcriptome Characterization of Canine Retina Using High-Throughput Sequencing. *Genetics Research International*, 2015, 638679.

Reese, M.G., Eeckman, F.H., Kulp, D., Haussler, D., (1997) Improved splice site detection in Genie. *Journal of computational biology : a journal of computational molecular cell biology*, 4(3):311-323.

RetNet, the Retinal Information Network, 2020. Available online: http://www.sph.uth.tmc.edu/RetNet/ (Accessed: 20 February 2020).

Riazuddin, S. A., Iqbal, M., Wang, Y., Masuda, T., Chen, Y., Bowne, S., Sullivan, L. S., Waseem, N. H., Bhattacharya, S., Daiger, S. P., Zhang, K., Khan, S. N., Riazuddin, S., Hejtmancik, J. F., Sieving, P. A., Zack, D. J. & Katsanis, N. (2010). A Splice-Site Mutation in a Retina-Specific Exon of BBS8 Causes Nonsyndromic Retinitis Pigmentosa. *American Journal of Human Genetics*, 86(5), 805-812.

Richardson, R., Tracey-White, D., Webster, A. & Moosajee, M. (2017). The zebrafish eye-a paradigm for investigating human ocular genetics. *Eye*, 31(1), 68-86.

Ritter, B., Philie, J., Girard, M., Tung, E. C., Blondeau, F. & McPherson, P. S. (2003). Identification of a family of endocytic proteins that define a new alpha-adaptin ear-binding motif. *EMBO Reports*, 4(11), 1089-1095.

Ritter, B., Blondeau, F., Denisov, A. Y., Gehring, K. & McPherson, P. S. (2004). Molecular mechanisms in clathrin-mediated membrane budding revealed through subcellular proteomics. *Biochemical Society Transactions*, 32(Pt 5), 769-773.

Ritter, B., Denisov, A. Y., Philie, J., Allaire, P. D., Legendre-Guillemin, V., Zylbergold, P., Gehring, K. & McPherson, P. S. (2007). The NECAP PHear domain increases clathrin accessory protein binding potential. *The EMBO Journal*, 26(18), 4066-4077.

Ritter, B., Murphy, S., Dokainish, H., Girard, M., Gudheti, M. V., Kozlov, G., Halin, M., Philie, J., Jorgensen, E. M., Gehring, K. & McPherson, P. S. (2013). NECAP 1 regulates AP-2 interactions to control vesicle size, number, and cargo during clathrin-mediated endocytosis. *PLoS Biol*, 11(10), e1001670.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24-26.

Rogers, A. J. & Weiss, S. T. (2017). Chapter 17 - Epidemiologic and Population Genetic Studies, in Robertson, D. & Williams, G. H. (eds), *Clinical and Translational Science (Second Edition)* Academic Press, 313-326.

Ropstad, E. O., Bjerkas, E. & Narfström, K. (2007a). Clinical findings in early onset cone-rod dystrophy in the Standard Wire-haired Dachshund. *Veterinary Ophthalmology*, 10(2), 69-75.

Ropstad, E. O., Bjerkas, E. & Narfström, K. (2007b). Electroretinographic findings in the Standard Wire Haired Dachshund with inherited early onset cone-rod dystrophy. *Documenta Ophthalmologica. Advances in Ophthalmology*, 114(1), 27-36.

Rosenbaum, J. L. & Witman, G. B. (2002). Intraflagellar transport. *Nature reviews. Molecular Cell Biology,* 3(11), 813-825.

Ross, J. W., Fernandez De Castro, J. P., Zhao, J., Samuel, M., Walters, E., Rios, C., Bray-Ward, P., Jones, B. W., Marc, R. E., Wang, W., Zhou, L., Noel, J. M., McCall, M. A., Demarco, P. J., Prather, R. S. & Kaplan, H. J. (2012). Generation of an Inbred Miniature Pig Model of Retinitis Pigmentosa. *Investigative Ophthalmology & Visual Science,* 53(1), 501-507.

Russell, S., Bennett, J., Wellman, J. A., Chung, D. C., Yu, Z.-F., Tillman, A., Wittes, J., Pappas, J., Elci, O., McCague, S., Cross, D., Marshall, K. A., Walshire, J., Kehoe, T. L., Reichert, H., Davis, M., Raffini, L., George, L. A., Hudson, F. P., Dingfield, L., Zhu, X., Haller, J. A., Sohn, E. H., Mahajan, V. B., Pfeifer, W., Weckmann, M., Johnson, C., Gewaily, D., Drack, A., Stone, E., Wachtel, K., Simonelli, F., Leroy, B. P., Wright, J. F., High, K. A. & Maguire, A. M. (2017). Efficacy and safety of voretigene neparvovec

(AAV2-hRPE65v2) in patients with RPE65 -mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *The Lancet*, 390(10097), 849-860.

Sanchis-Juan, A., Stephens, J., French, C. E., Gleadall, N., Mégy, K., Penkett, C., Shamardina, O., Stirrups, K., Delon, I., Dewhurst, E., Dolling, H., Erwood, M., Grozeva, D., Stefanucci, L., Arno, G., Webster, A. R., Cole, T., Austin, T., Branco, R. G., Ouwehand, W. H., Raymond, F. L. & Carss, K. J. (2018). Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Medicine*, 10(1).

Sato, S., Peshenko, I. V., Olshevskaya, E. V., Kefalov, V. J. & Dizhoor, A. M. (2018). GUCY2D Cone–Rod Dystrophy-6 Is a "Phototransduction Disease" Triggered by Abnormal Calcium Feedback on Retinal Membrane Guanylyl Cyclase 1. *The Journal of Neuroscience*, 38(12), 2990-3000.

Savolainen, P., Zhang, Y. P., Luo, J., Lundeberg, J. & Leitner, T. (2002). Genetic evidence for an East Asian origin of domestic dogs. *Science*, 298(5598), 1610-1613.

Scheidecker, S., Etard, C., Pierce, N. W., Geoffroy, V., Schaefer, E., Muller, J., Chennen, K., Flori, E., Pelletier, V., Poch, O., Marion, V., Stoetzel, C., Strähle, U., Nachury, M. V. & Dollfus, H. (2014). Exome sequencing of Bardet–Biedl syndrome patient identifies a null mutation in the BBSome subunit BBIP1(BBS18). *Journal of Medical Genetics*, 51(2), 132-136.

Scherf, M., Klingenhoff, A. & Werner, T. (2000). Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *Journal of Molecular Biology,*, 297(3), 599-606.

Seddon, J. M., Hampson, E. C., Smith, R. I. & Hughes, I. P. (2006). Genetic heterogeneity of day blindness in Alaskan Malamutes. *Animal Genetics*, 37(4), 407-410.

Seo, S., Guo, D. F., Bugge, K., Morgan, D. A., Rahmouni, K. & Sheffield, V. C. (2009). Requirement of Bardet-Biedl syndrome proteins for leptin receptor signaling. *Human Molecular Genetics*, 18(7), 1323-1331.

Seo, S., Baye, L. M., Schulz, N. P., Beck, J. S., Zhang, Q., Slusarski, D. C. & Sheffield, V. C. (2010). BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 107(4), 1488-1493.

Seo, S., Zhang, Q., Bugge, K., Breslow, D. K., Searby, C. C., Nachury, M. V. & Sheffield, V. C. (2011). A Novel Protein LZTFL1 Regulates Ciliary Trafficking of the BBSome and Smoothened. *PLoS Genetics,* 7(11), e1002358.

Seo, S. & Datta, P. (2017). Photoreceptor outer segment as a sink for membrane proteins: hypothesis and implications in retinal ciliopathies. *Human Molecular Genetics*, 26(R1), R75-R82.

Shevach, E., Ali, M., Mizrahi-Meissonnier, L., McKibbin, M., El-Asrag, M., Watson, C. M., Inglehearn, C. F., Ben-Yosef, T., Blumenfeld, A., Jalas, C., Banin, E. & Sharon, D. (2015). Association between missense mutations in the BBS2 gene and nonsyndromic retinitis pigmentosa. *JAMA Ophthalmology*, 133(3), 312-318.

Sidjanin, D. J., Lowe, J. K., McElwee, J. L., Milne, B. S., Phippen, T. M., Sargan, D. R., Aguirre, G. D., Acland, G. M. & Ostrander, E. A. (2002). Canine CNGB3 mutations establish cone degeneration as orthologous to the human achromatopsia locus ACHM3. *Human Molecular Genetics*, 11(16), 1823-1833.

Siemiatkowska, A. M., Collin, R. W. J., Den Hollander, A. I. & Cremers, F. P. M. (2014). Genomic Approaches For the Discovery of Genes Mutated in Inherited Retinal Degeneration. *Cold Spring Harbor Perspectives in Medicine,* 4(8), a017137.

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40(Web Server issue), W452-457.

Simonelli, F., Maguire, A. M., Testa, F., Pierce, E. A., Mingozzi, F., Bennicelli, J. L., Rossi, S., Marshall, K., Banfi, S., Surace, E. M., Sun, J., Redmond, T. M., Zhu, X., Shindler, K. S., Ying, G. S., Ziviello, C., Acerra, C., Wright, J. F., McDonnell, J. W., High, K. A., Bennett, J. & Auricchio, A. (2010). Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. *Molecular Therapy : the Journal of the American Society of Gene Therapy*, 18(3), 643-650.

Simó, R., Villarroel, M., Corraliza, L., Hernández, C. & Garcia-Ramírez, M. (2010). The Retinal Pigment Epithelium: Something More than a Constituent of the Blood-Retinal Barrier—Implications for the Pathogenesis of Diabetic Retinopathy. *Journal of Biomedicine and Biotechnology*, 2010, 1-15.

Singh, A. K., Khare, P., Obaid, A., Conlon, K. P., Basrur, V., Depinho, R. A. & Venuprasad, K. (2018). SUMOylation of ROR-γt inhibits IL-17 expression and inflammation via HDAC2. *Nature Communications*, 9(1).

Slavotinek, A. M., Stone, E. M., Mykytyn, K., Heckenlively, J. R., Green, J. S., Heon, E., Musarella, M. A., Parfrey, P. S., Sheffield, V. C. & Biesecker, L. G. (2000). Mutations in MKKS cause Bardet-Biedl syndrome. *Nature Genetics*, 26(1), 15-16.

Smith, B. F., Yue, Y., Woods, P. R., Kornegay, J. N., Shin, J. H., Williams, R. R. & Duan, D. (2011). An intronic LINE-1 element insertion in the dystrophin gene aborts dystrophin expression and results in Duchenne-like muscular dystrophy in the corgi breed. *Laboratory Investigation; A Journal of Technical Methods and Pathology*, 91(2), 216-231.

Spektor, A., Tsang, W. Y., Khoo, D. & Dynlacht, B. D. (2007). Cep97 and CP110 suppress a cilia assembly program. *Cell*, 130(4), 678-690.

Staden, R., Beal, K. F. & Bonfield, J. K. (2000). The Staden package, 1998. *Methods in Molecular Biology*, 132, 115-130.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M. & Lancet, D. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics*, 54, 1.30.1-1.30.33.

Stiebel-Kalish, H., Reich, E., Rainy, N., Vatine, G., Nisgav, Y., Tovar, A., Gothilf, Y. & Bach, M. (2012). Gucy2f zebrafish knockdown--a model for Gucy2d-related leber congenital amaurosis. *European Journal of Human Genetics : EJHG*, 20(8), 884-889.

Stoetzel, C., Laurier, V., Davis, E. E., Muller, J., Rix, S., Badano, J. L., Leitch, C. C., Salem, N., Chouery, E., Corbani, S., Jalk, N., Vicaire, S., Sarda, P., Hamel, C., Lacombe, D., Holder, M., Odent, S., Holder, S., Brooks, A. S., Elcioglu, N. H., Silva, E. D., Rossillion, B., Sigaudy, S., de Ravel, T. J. L., Lewis, R. A., Leheup, B., Verloes, A., Amati-Bonneau, P., Mégarbané, A., Poch, O., Bonneau, D., Beales, P. L., Mandel, J.-L., Katsanis, N. & Dollfus, H. (2006). BBS10 encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus. *Nature Genetics*, 38(5), 521-524.

Stoetzel, C., Muller, J., Laurier, V., Davis, E. E., Zaghloul, N. A., Vicaire, S., Jacquelin, C., Plewniak, F., Leitch, C. C., Sarda, P., Hamel, C., De Ravel, T. J. L., Lewis, R. A., Friederich, E., Thibault, C., Danse, J.-M., Verloes, A., Bonneau, D., Katsanis, N., Poch, O., Mandel, J.-L. & Dollfus, H. (2007) Identification of a Novel BBS Gene (BBS12). Highlights the Major Role of a Vertebrate-Specific Branch of Chaperonin-Related Proteins in Bardet-Biedl Syndrome. *American Journal of Human Genetics*, 80(1), 1-11.

Strom, T. M., Nyakatura, G., Apfelstedt-Sylla, E., Hellebrand, H., Lorenz, B., Weber, B. H., Wutz, K., Gutwillinger, N., Ruther, K., Drescher, B., Sauer, C., Zrenner, E., Meitinger, T., Rosenthal, A. & Meindl, A. (1998). An L-type calcium-channel gene mutated in incomplete X-linked congenital stationary night blindness. *Nature Genetics*, 19(3), 260-263.

Suber, M. L., Pittler, S. J., Qin, N., Wright, G. C., Holcombe, V., Lee, R. H., Craft, C. M., Lolley, R. N., Baehr, W. & Hurwitz, R. L. (1993). Irish setter dogs affected with rod/cone dysplasia contain a nonsense mutation in the rod cGMP phosphodiesterase beta-subunit gene. *Proceedings of the National Academy of Sciences of the United States of America*, 90(9), 3968-3972.

Sullivan, L. S., Bowne, S. J., Koboldt, D. C., Cadena, E. L., Heckenlively, J. R., Branham, K. E., Wheaton, D. H., Jones, K. D., Ruiz, R. S., Pennesi, M. E., Yang, P., Davis-Boozer, D., Northrup, H., Gurevich, V. V., Chen, R., Xu, M., Li, Y., Birch, D. G. & Daiger, S. P. (2017). A Novel Dominant Mutation in SAG, the Arrestin-1 Gene, Is a Common Cause of Retinitis Pigmentosa in Hispanic Families in the Southwestern United States. *Investigative Opthalmology & Visual Science*, 58(5), 2774.

Sutter, N. B., Eberle, M. A., Parker, H. G., Pullar, B. J., Kirkness, E. F., Kruglyak, L. & Ostrander, E. A. (2004). Extensive and breed-specific linkage disequilibrium in Canis familiaris. *Genome Research*, 14(12), 2388-2396.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J. & Christian (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1), D607-D613.

Takei, K. & Haucke, V. (2001). Clathrin-mediated endocytosis: membrane factors pull the trigger. *Trends in Cell Biology*, 11(9), 385-391.

Tanaka, N., Dutrow, E. V., Miyadera, K., Delemotte, L., MacDermaid, C. M., Reinstein, S. L., Crumley, W. R., Dixon, C. J., Casal, M. L., Klein, M. L., Aguirre, G. D., Tanaka, J. C. & Guziewicz, K. E. (2015). Canine CNGA3 Gene Mutations Provide Novel Insights into Human Achromatopsia-Associated Channelopathies and Treatment. *PLoS One*, 10(9), e0138943.

The Kennel Club UK (2019a). *Summary of registrations for all groups*. Available online: http://www.thekennelclub.org.uk/media/128966/group stats.pdf (Accessed: 26 March 2020).

The Kennel Club UK (2019b). *Breed Standard*. Available online: https://www.thekennelclub.org.uk/services/public/breed/standard.aspx?id=5140 (Accessed: 26 March 2020).

The Kennel Club UK (2015). *Population analysis of the Shetland Sheepdog breed*. Available online: https://www.thekennelclub.org.uk/media/686496/shetland sheepdog.pdf (Accessed: 26 March 2020).

The Kennel Club UK (2020). *Comparative tables of registrations for the years 2010- 2019 inclusive*. Available online: https://www.thekennelclub.org.uk/media/129025/10yrstatsworking.pdf (Accessed: 26 March 2020).

Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178-192.

Tieder, M., Levy, M., Gubler, M. C., Gagnadoux, M. F. & Broyer, M. (1982). Renal abnormalities in the Bardet-Biedl syndrome. *The International Journal of Pediatric Nephrology*, 3(3), 199-203.

Tobin, J. L., Di Franco, M., Eichers, E., May-Simera, H., Garcia, M., Yan, J., Quinlan, R., Justice, M. J., Hennekam, R. C., Briscoe, J., Tada, M., Mayor, R., Burns, A. J., Lupski, J. R., Hammond, P. & Beales, P. L. (2008). Inhibition of neural crest migration underlies craniofacial dysmorphology and Hirschsprung's disease in Bardet-Biedl syndrome. *Proceedings of the National Academy of Sciences of the United States of America,* 105(18), 6714-6719.

Travis, G. H. (1998). Mechanisms of Cell Death in the Inherited Retinal Degenerations. *The American Journal of Human Genetics*, 62(3), 503-508.

Travis, G. H., Golczak, M., Moise, A. R., & Palczewski, K. (2007). Diseases caused by defects in the visual cycle: retinoids as potential therapeutic agents. *Annual review of pharmacology and toxicology*, 47, 469–512.

Turkdogan, D., Usluer, S., Akalin, F., Agyuz, U. & Aslan, E. S. (2017). Familial early infantile epileptic encephalopathy and cardiac conduction disorder: A rare cause of SUDEP in infancy. *Seizure*, 50, 171-172.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J. & Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220), 1260419.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115.

Valente, E. M., Silhavy, J. L., Brancati, F., Barrano, G., Krishnaswami, S. R., Castori, M., Lancaster, M. A., Boltshauser, E., Boccone, L., Al-Gazali, L., Fazzi, E., Signorini, S., Louie, C. M., Bellacchio, E., International Joubert Syndrome Related Disorders Study, G., Bertini, E., Dallapiccola, B. & Gleeson,

J. G. (2006). Mutations in CEP290, which encodes a centrosomal protein, cause pleiotropic forms of Joubert syndrome. *Nature Genetics*, 38(6), 623-5.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S. & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43, 11.10.1-11.10.33.

Veleri, S., Lazar, C. H., Chang, B., Sieving, P. A., Banin, E. & Swaroop, A. (2015). Biology and therapy of inherited retinal degenerative disease: insights from mouse models. *Disease Models & Mechanisms*, 8(2), 109-129.

Veleri, S., Nellissery, J., Mishra, B., Manjunath, S. H., Brooks, M. J., Dong, L., Nagashima, K., Qian, H., Gao, C., Sergeev, Y. V., Huang, X. F., Qu, J., Lu, F., Cideciyan, A. V., Li, T., Jin, Z. B., Fariss, R. N., Ratnapriya, R., Jacobson, S. G. & Swaroop, A. (2017). REEP6 mediates trafficking of a subset of Clathrin-coated vesicles and is critical for rod photoreceptor function and survival. *Human Molecular Genetics*, 26(12), 2218-2230.

Verbakel, S. K., van Huet, R. A. C., Boon, C. J. F., den Hollander, A. I., Collin, R. W. J., Klaver, C. C. W., Hoyng, C. B., Roepman, R. & Klevering, B. J. (2018). Non-syndromic retinitis pigmentosa. *Progress in Retinal Eye Research*, 66, 157-186.

Veske, A., Nilsson, S. E., Narfström, K. & Gal, A. (1999). Retinal dystrophy of Swedish briard/briard-beagle dogs is due to a 4-bp deletion in RPE65. *Genomics*, 57(1), 57-61.

Vila, C., Savolainen, P., Maldonado, J. E., Amorim, I. R., Rice, J. E., Honeycutt, R. L., Crandall, K. A., Lundeberg, J. & Wayne, R. K. (1997). Multiple and ancient origins of the domestic dog. *Science*, 276(5319), 1687-1689.

Vilboux, T., Chaudieu, G., Jeannin, P., Delattre, D., Hedan, B., Bourgain, C., Queney, G., Galibert, F., Thomas, A., & André, C. (2008). *Progressive retinal atrophy in the Border Collie: a new XLPRA*. BMC veterinary research, 4, 10.

Vollrath, D., Feng, W., Duncan, J. L., Yasumura, D., D'Cruz, P. M., Chappelow, A., Matthes, M. T., Kay, M. A. & LaVail, M. M. (2001). Correction of the retinal dystrophy phenotype of the RCS rat by viral gene transfer of Mertk. *Proceedings of the National Academy of Sciences of the United States of America*, 98(22), 12584-12589.

Waters, A. M. & Beales, P. L. (2011). Ciliopathies: an expanding disease spectrum. *Pediatric Nephrology*, 26(7), 1039-1056.

Webb, T. R., Parfitt, D. A., Gardner, J. C., Martinez, A., Bevilacqua, D., Davidson, A. E., Zito, I., Thiselton, D. L., Ressa, J. H., Apergi, M., Schwarz, N., Kanuga, N., Michaelides, M., Cheetham, M. E., Gorin, M. B. & Hardcastle, A. J. (2012). Deep intronic mutation in OFD1, identified by targeted genomic next-generation sequencing, causes a severe form of X-linked retinitis pigmentosa (RP23). *Human Molecular Genetics*, 21(16), 3647-3654.

Weleber, R. G. & Gregory-Evans, K. (2006). Chapter 17 - Retinitis Pigmentosa and Allied Disorders, in Ryan, S. J., Hinton, D. R., Schachat, A. P. & Wilkinson, C. P. (eds), *Retina (Fourth Edition)*. Edinburgh: Mosby, 395-498.

Wiik, A. C., Wade, C., Biagi, T., Ropstad, E. O., Bjerkas, E., Lindblad-Toh, K. & Lingaas, F. (2008). A deletion in nephronophthisis 4 (NPHP4) is associated with recessive cone-rod dystrophy in standard wire-haired dachshund. *Genome Research*, 18(9), 1415-1421.

Wiik, A. C., Ropstad, E. O., Ekesten, B., Karlstam, L., Wade, C. M. & Lingaas, F. (2015). Progressive retinal atrophy in Shetland sheepdog is associated with a mutation in the CNGA1 gene. *Animal Genetics*, 46(5), 515-521.

Won, J., Gifford, E., Smith, R. S., Yi, H., Ferreira, P. A., Hicks, W. L., Li, T., Naggert, J. K. & Nishina, P. M. (2009). RPGRIP1 is essential for normal rod photoreceptor outer segment elaboration and morphogenesis. *Human Molecular Genetics*, 18(22), 4329-4339.

Wrobel, A. G., Kadlecova, Z., Kamenicky, J., Yang, J.-C., Herrmann, T., Kelly, B. T., McCoy, A. J., Evans, P. R., Martin, S., Müller, S., Sroubek, F., Neuhaus, D., Höning, S. & Owen, D. J. (2019). Temporal Ordering in Endocytic Clathrin-Coated Vesicle Formation via AP2 Phosphorylation. *Developmental Cell*, 50(4), 494-508.e11.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T. W., Li, Y., Xu, X., Wong, G. K. & Wang, J. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, 30(12), 1660-1666.

Xing, D.-J., Zhang, H.-X., Huang, N., Wu, K.-C., Huang, X.-F., Huang, F., Tong, Y., Pang, C.-P., Qu, J. & Jin, Z.-B. (2014). Comprehensive Molecular Diagnosis of Bardet-Biedl Syndrome by High-Throughput Targeted Exome Sequencing. *PloS one,* 9(3), e90599.

Xiong, B. & Bellen, H. J. (2013). Rhodopsin homeostasis and retinal degeneration: lessons from the fly. *Trends in Neuroscience*, 36(11), 652-660.

Xu, J., Dodd, R. L., Makino, C. L., Simon, M. I., Baylor, D. A. & Chen, J. (1997). Prolonged photoresponses in transgenic mouse rods lacking arrestin. *Nature*, 389(6650), 505-509.

Xu, M., Yang, L., Wang, F., Li, H., Wang, X., Wang, W., Ge, Z., Wang, K., Zhao, L., Li, H., Li, Y., Sui, R. & Chen, R. (2015). Mutations in human IFT140 cause non-syndromic retinal degeneration. *Human Genetics* 134(10), 1069-1078.

Yang, F., O'Brien, P. C., Milne, B. S., Graphodatsky, A. S., Solanky, N., Trifonov, V., Rens, W., Sargan, D., & Ferguson-Smith, M. A. (1999). A complete comparative chromosome map for the dog, red fox, and human and its integration with canine genetic maps. *Genomics*, 62(2), 189–202.

Yeh, C. Y., Goldstein, O., Kukekova, A. V., Holley, D., Knollinger, A. M., Huson, H. J., Pearce-Kelling, S. E., Acland, G. M. & Komaromy, A. M. (2013). Genomic deletion of CNGB3 is identical by descent in multiple canine breeds and causes achromatopsia. *BMC Genetics*, 14, 27.

Yıldız Bölükbaşı, E., Mumtaz, S., Afzal, M., Woehlbier, U., Malik, S. & Tolun, A. (2018). Homozygous mutation in CEP19, a gene mutated in morbid obesity, in Bardet-Biedl syndrome with predominant postaxial polydactyly. *Journal of Medical Genetics*, 55(3), 189-197.

Young, R. W. (1967). The renewal of photoreceptor cell outer segments. *The Journal of Cell Biology*, 33(1), 61-72.

Zangerl, B., Goldstein, O., Philp, A. R., Lindauer, S. J., Pearce-Kelling, S. E., Mullins, R. F., Graphodatsky, A. S., Ripoll, D., Felix, J. S., Stone, E. M., Acland, G. M. & Aguirre, G. D. (2006). Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. *Genomics*, 88(5), 551-563.

Zangerl, B., Wickstrom, K., Slavik, J., Lindauer, S. J., Ahonen, S., Schelling, C., Lohi, H., Guziewicz, K. E. & Aguirre, G. D. (2010). Assessment of canine BEST1 variations identifies new mutations and establishes an independent bestrophinopathy model (cmr3). *Molecular Vision*, 16, 2791-804.

Zeiss, C. J., Ray, K., Acland, G. M., & Aguirre, G. D. (2000). Mapping of X-linked progressive retinal atrophy (XLPRA), the canine homolog of retinitis pigmentosa 3 (RP3). *Human Molecular Genetics*, 9(4), 531–537.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., To, J. K., Laird, M. R., Lavidas, I., Liu, Z., Loveland, J. E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D. N., Newman, V., Nuhn, M., Ogeh, D., Ong, C. K., Parker, A., Patricio, M., Riat, H. S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S. E., Kostadima, M., Langridge, N., Martin, F. J., Muffato, M., Perry, E., Ruffier, M., Staines, D. M., Trevanion, S. J., Aken, B. L., Cunningham, F., Yates, A. & Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, 46(D1), D754-D761.

Zhang, Q., Acland, G. M., Parshall, C. J., Haskell, J., Ray, K. & Aguirre, G. D. (1998). Characterization of canine photoreceptor phosducin cDNA and identification of a sequence variant in dogs with photoreceptor dysplasia. *Gene*, 215(2), 231-239.

Zhang, Q., Acland, G. M., Zangerl, B., Johnson, J. L., Mao, Z., Zeiss, C. J., Ostrander, E. A., & Aguirre, G. D. (2001). Fine mapping of canine XLPRA establishes homology of the human and canine RP3 intervals. *Investigative Ophthalmology & Visual Science*, 42(11), 2466–2471.

Zhang, Q., Acland, G. M., Wu, W. X., Johnson, J. L., Pearce-Kelling, S., Tulloch, B., Vervoort, R., Wright, A. F. & Aguirre, G. D. (2002). Different RPGR exon ORF15 mutations in Canids provide insights into photoreceptor cell degeneration. *Human Molecular Genetics*, 11(9), 993-1003.

Zhang, Q., Yu, D., Seo, S., Stone, E. M. & Sheffield, V. C. (2012). Intrinsic Protein-Protein Interaction-mediated and Chaperonin-assisted Sequential Assembly of Stable Bardet-Biedl Syndrome Protein Complex, the BBSome, 287(24), 20625-20635.

Zhang, F., Meng, G. & Strober, W. (2008). Interactions among the transcription factors Runx1, RORgammat and Foxp3 regulate the differentiation of interleukin 17-producing T cells. *Nature Immunology*, 9(11), 1297-1306.

Zhao, Y., Hong, D. H., Pawlyk, B., Yue, G., Adamian, M., Grynberg, M., Godzik, A. & Li, T. (2003). The retinitis pigmentosa GTPase regulator (RPGR)- interacting protein: subserving RPGR function and participating in disk morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), 3965-3970.

Zuchner, S., Dallman, J., Wen, R., Beecham, G., Naj, A., Farooq, A., Kohli, M. A., Whitehead, P. L., Hulme, W., Konidari, I., Edwards, Y. J., Cai, G., Peter, I., Seo, D., Buxbaum, J. D., Haines, J. L., Blanton, S., Young, J., Alfonso, E., Vance, J. M., Lam, B. L. & Pericak-Vance, M. A. (2011). Whole-exome sequencing links a variant in DHDDS to retinitis pigmentosa. *American Journal of Human Genetics*, 88(2), 201-206.